Measurement of the lifetimes of B meson mass eigenstates

by

Konstantin Anikeev

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2004

Author .....
Department of Physics
September 28, 2004

Certified by .....
Christoph M. E. Paus
Associate Professor
Thesis Supervisor

Accepted by .....
Thomas J. Greytak
Associate Department Head for Education, Professor

Measurement of the lifetimes of B meson mass eigenstates

by

Konstantin Anikeev

Submitted to the Department of Physics
on September 28, 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

In this dissertation, we present the results of the average lifetime measurements in $B_u \to J/\psi K^+$, $B_d \to J/\psi K^{*0}$, and $B_s \to J/\psi\phi$ decays, as well as the results of a time-dependent angular analysis of $B_d \to J/\psi K^{*0}$ and $B_s \to J/\psi\phi$ decays.

The time-dependent angular analysis of $B_s \to J/\psi\phi$ allows the extraction of separate lifetimes for the mass eigenstates, B_s^L and B_s^H, in the B_s–$\overline{B}_s$ system. The two lifetimes are combined to extract the width difference, $\Delta\Gamma_s \equiv \Gamma_s^L - \Gamma_s^H = 1/\tau_{B_s^L} - 1/\tau_{B_s^H}$, between the two eigenstates. This quantity is extremely useful for over-constraining the Unitarity Triangle and thereby inferring if CP violation in the quark sector is entirely explained within the Standard Model.

From $258 \pm 15\,\mathrm{pb}^{-1}$ of data collected by CDF up to February 13, 2004, we obtain a wealth of results, most importantly:

$$\begin{aligned}
\tau_{B_u} &= (1.659 \pm 0.033\,^{+0.007}_{-0.008})\,\mathrm{ps} \\
\tau_{B_d} &= (1.549 \pm 0.051\,^{+0.007}_{-0.008})\,\mathrm{ps} \\
\tau_{B_s^L} &= (1.05\,^{+0.16}_{-0.13} \pm 0.02)\,\mathrm{ps} \\
\tau_{B_s^H} &= (2.07\,^{+0.58}_{-0.46} \pm 0.03)\,\mathrm{ps} \\
\Delta\Gamma_s &= (0.46 \pm 0.18 \pm 0.01)\,\mathrm{ps}^{-1}
\end{aligned}$$

Several of the results obtained here are currently the best of their kind, and in some cases they are unique measurements. We devote special attention to these, in particular $\Delta\Gamma_s$.

Thesis Supervisor: Christoph M. E. Paus
Title: Associate Professor

... to that girl

Acknowledgment

This dissertation represents my most important physics accomplishment to date. However, it would have been of lower quality or, perhaps would not have happened at all, without help and inspiration provided by many people. I treat the current section as the dearest of all, because it is in some way my first real chance to say "thank you" to those, who contributed to my progress, often unknowingly or indirectly. It may be a little longer than a casual reader would expect the acknowledgment section to be — the reason being that I was blessed to meet not just a few, but many wonderful people. Most of them I would really like to mention here.

First, I thank my advisor Christoph Paus. Looking back I see that he was the right person to make me go through: ambitious, and at times tough, but resourceful. I doubt many students could brag about their advisor helping them to get to terms with `PERL` scripts, `C++`, data handling issues, and other "technical stuff". In modern particle physics however, getting the "technical stuff" right is more than half of the job; and this is where Christoph was instrumental like no one else. In doing "real physics" Christoph stands out as a scientist who can clearly define a problem and often come up with an efficient way of attacking it. Whenever time allowed, he has been open to questions. Some of the questions led to interesting discussions, or even arguments, from which I benefited substantially. I also appreciate his ability to trust students to do analysis independently, giving them time to think and come up with their own decisions.

A significant amount of expertise came from my colleagues, young MITers: my office mate Ivan Furic[1], Andreas Korn[2], our ever cool postdoc and now research scientist Ilya Kravchenko, my Time-of-Flight buddy Mike Mulhearn, Boris Iyutin, Alberto Belloni and Stefan Gromoll. Another MIT student, Alexandr Rakitin, deserves additional recognition, because he endured[3] being my roommate for many of these years.

More MIT people who I absolutely must mention are Steve Tether, Steve Pavlon, Sham Sumorok, and Frank Würthwein. They were an invaluable source of assistance be it computer problems, technical issues in developing Time-of-Flight system, B physics theory or just textual issues when writing papers. Big and very special thanks to Gerry Bauer for always finding time to discuss those numerous "philosophical questions". He is one nice person to talk to.

I express my gratitude to Matt Herndon, Tom LeCompte, Ting Miao, Mary Bishai, Kevin Pitts, Marjorie Shapiro and Stefano Giagu, who have been at various times conveners of the CDF J/ψ and B groups, for their expertise, guidance and support.

I thank Rich Stanek, who exemplifies Fermilab's top notch engineers and technicians. We worked on building CDF Time-of-Flight system, with Rich leading the design part of the project. Nevertheless, he had also gone to the pit and worked as much as others transforming drawings into tangible detector parts.

I remember with warmth Jon Miller, who showed me around CDF and had helped

[1] VeRRy impRRessive misteRR Bond.

[2] I thought I was the unhappiest man on Earth, but sometimes he proved me wrong :o) .

[3] Or was it the other way around? Let history decide...

me throughout the first year, as well as Tushar Shah for his jokes, math puzzles and monitoring my "happiness level" through the beginning years. Peggy Berkovitz was a great help during those times and by far the best administrative assistant I have ever had to do business with.

I thank Robyn Madrak and Luis Flores Castillo, who had worked on related analyses, for sharing their know-how. I also thank CDF/Yale folks: Ke Li, Colin Gay, and Michael Schmidt, as well as Denis Bernard from BaBar for illuminating discussions, many of which have made their way into this thesis in one form or another.

I am proud and grateful to have been a member of the "CDF Run Too!!" crew led by Greg Feild and tireless Eugene "JJ" Schmidt. By being part of the "CDF Run Too!!" not only did I get a chance to participate in exciting races, but also I got to know some of my CDF colleagues outside the rather stressful work day environment.

I am equally grateful to almost all of the people I played with in the Fermilab Soccer League, which often served as a much needed relief valve. Rumor has it that I played tough and might have hurt other players. Without lying to myself, I can say that I have always made a conscious effort to stay within the rules, if nothing else but for the reason that the game is not the game if the rules are broken. I do apologize however, for any pains I might have inflicted – it has *never* been intentional.

Paul DiTuro and Steve Nahn, who also happened to be involved in running and soccer, should not be forgotten. Their honesty and enthusiasm influenced me and, I believe, made my CDF experience a much brighter one.

I was lucky to be welcomed – and in certain sense fostered – by a number of Russian families here, at Fermilab. Families of Gotra, Kashikhin, Los, Nomerotsky, Novitsky, Orlov, Peryshkin, Podstavkov, Prokofiev, Shemyakin, and Sidorov did make me feel at home and saved me from becoming a miserable workaholic[4] by introducing me to Family Soccer, taking me to numerous camping, rafting, and mountain hiking trips, inviting me to various parties and celebrations and offering the feel of being at home in many other ways.

Back in Russia, I thank my school friends Pyotr Dolgopolov, Denis Snitko, Andrey Moskalev, and Sergey Kalugin as well as my cousin Sergey Pchelincev. I will treasure the experiences we have gone through for the rest of my life. I also thank my wrestling coach Victor Minaev and ship modeling instructor Sergey Drugachenok for everything I learned from them.

There are no words to describe how indebted I am to my very best friends and fellow FizTeh students: Sergey Dzhosyuk, Alexey Sedov, Nikolay Slobozhanov, Pyotr Puzynia, Marat Rvachev, Misha Tolstikov, Vadim Stakhursky, Victor Dergunov. Several years ago I was not sure if I could stay in physics and continue doing what I like and consider important, purely for economic reasons. It was them who opened my eyes to the possibility of moving across the ocean and making the dream of working at the world's leading particle physics laboratory come true. Largely due to their effort we were able to put together a study group to adequately prepare for the exams and the graduate school application process. My friends were also the people who have

[4]There is still an ongoing argument whether I could actually get the disease :o) .

helped me understand and change my once not-so-good personality. Thank you guys for everything. If you ever need my help — I will be there.

I thank my family, near and far, for supporting me during student years in Russia. Those years were full of hardship – we know it all too well – but because of you and my friends these had also become my best years.

Above all, I thank Mom and Dad, for raising me the way I am, for spending countless hours making sure that I get the best education, while my body does not get weak and is ready to withstand whatever life brings on. Mom and Dad, I love you.

Contents

1 Introduction **13**

1.1 Weak interaction and the CKM description of the flavor sector 13
1.1.1 Historical development 14
1.1.2 Weak Lagrangian: the origin of the CKM Matrix 16
1.1.3 CKM Matrix and the Unitarity Triangle 18
1.2 Brief theoretical overview of B meson lifetimes 22
1.2.1 B lifetimes: the spectator model and beyond 23
1.2.2 HQE review 25
1.3 Mixing phenomenon in neutral B mesons 28
1.3.1 Quantum-mechanical treatment 28
1.3.2 Phenomenology 30
1.4 Angular correlations in the decays of neutral B mesons 33
1.4.1 Transversity basis and transversity variables 33
1.4.2 Time-dependent angular distributions in the $B_d \to J/\psi K^{*0}$ and $B_s \to J/\psi\phi$ decays 34
1.5 Analysis overview 37
1.5.1 B lifetimes 38
1.5.2 Angular decay amplitudes in the $P \to VV$ B meson decays . . 39
1.5.3 The status of and considerations for $\Delta\Gamma_s/\Gamma_s$ 40
1.5.4 Analysis strategy 42

2 Experimental apparatus and getting the data **45**

2.1 Tevatron 45
2.1.1 Proton source: Cockroft-Walton chamber, Linac, and Booster 45
2.1.2 Main Injector 47
2.1.3 Antiproton source: Debuncher, Accumulator, and Recycler . . 48
2.1.4 Tevatron – the global picture 50
2.2 The CDF II detector 52
2.2.1 Tracking and vertexing 54
2.2.2 Muon detection 64
2.2.3 Other systems 69
2.3 Triggers and data acquisition 72
2.3.1 Level-1 73
2.3.2 Level-2 76
2.3.3 Event Builder and Level-3 trigger 76

2.4 Detector operation and off-line data handling 77

3 Analysis of the data **79**
3.1 Data and candidate selection 80
3.1.1 Treatment of tracks 81
3.1.2 Treatment of muons 82
3.1.3 J/ψ, K^{*0} and ϕ selection 82
3.1.4 B meson selection 84
3.1.5 Other selection caveats 84
3.1.6 Nominal cuts 86
3.2 Sample composition and fitting technique 86
3.2.1 Fit variables 88
3.2.2 PDF for the mass (m) 90
3.2.3 PDF for the proper decay length (ct) 92
3.2.4 PDF for the transversity ($\vec{\omega}$) 94
3.2.5 Fit functions used in the analysis 105
3.3 Fit details and results 107
3.3.1 Fitting for the average lifetime 107
3.3.2 Time-dependent angular analysis 110
3.4 Systematic uncertainties 117
3.4.1 Selection 117
3.4.2 SVX alignment 119
3.4.3 Fit model 121
3.4.4 Residual disagreement between data and Monte-Carlo 127
3.4.5 B_d candidates with K $\leftrightarrow \pi$ mass mis-assignment 128
3.4.6 Residual $B_d \leftrightarrow B_s$ cross-feed 128
3.4.7 Unequal number of B_s and $\overline{B}_s$ 130
3.4.8 Non-resonant contribution to $\mu^+\mu^- K^+\pi^-$ final state 131
3.4.9 Other sources 131
3.4.10 Summary 132
3.5 Cross checks 133
3.5.1 Fitter tests 134
3.5.2 CP-odd fraction versus the PDL cut value 135
3.5.3 Constraining the B_s amplitudes 136
3.5.4 Single lifetime fit in the B_s sample 136
3.5.5 Estimated number of candidates at large PDL 137
3.6 Summary and discussion 138
3.6.1 Final results 138
3.6.2 Comparisons to other results 139
3.6.3 What are the betting odds? 144
3.6.4 Discussion 145
3.6.5 Implications 146
3.6.6 Conclusion 146

A Single lifetime likelihood fit **149**

B Good run definition **151**

C Realistic Monte-Carlo simulation **153**

D Cut optimization **161**
D.1 Cut optimization for the average lifetime measurement 161
D.2 Other considerations in the choice of cuts 164
D.3 P_T cut optimization for $\Delta\Gamma_s/\Gamma_s$ measurement 167

E $B_d \to J/\psi K^{*0}$ and $B_s \to J/\psi\phi$ cross-feed **169**

F Eliminating events with large negative ct **175**

G Calculating convolutions **177**

H Treating $B_u \to J/\psi\pi^+$ in B_u sample **179**

I Treating $K \leftrightarrow \pi$ self-reflection in B_d sample **181**

J Fitting for angular decay amplitudes or time-integrated fractions **185**

K Calculation of ξ_i **189**
K.1 ξ_i for B_s 189
K.2 ξ_i^{un} and ξ_i^{sw} for B_d 189
K.3 Summary 191

L Goodness of fit estimate **193**

M Projecting fit results on the angular distributions **195**

N Useful transformations **197**
N.1 Variable definitions 197
N.2 Relations 198
N.3 Summary 199

Chapter 1

Introduction

B mesons are bound states of an anti-b quark and a lighter quark. They are commonly denoted B_q or $\bar{b}q$, where q stands for a u, d, s, or c quark. B mesons are the only mesons which contain a quark of the third generation. The t quark has a mass much larger than the mass of the W boson and therefore decays before hadronization can occur.

The binding between quarks in a B meson is provided by the strong force. At the same time, ground state B mesons can only decay via the weak interaction, in most cases by means of the $b \to Wq$ sub-process, where q is either a c or a u quark and the emitted W boson is highly virtual. Studying B mesons gives physicists a handle on otherwise inaccessible elements of the Cabibbo-Kobayashi-Maskawa (CKM) matrix V_{cb}, V_{ub}, V_{td}, and V_{ts}. Precise determination of the CKM Matrix is one of the highest priorities of contemporary high energy physics, which makes measurements on B mesons extremely interesting and exciting.

In what follows we discuss the weak flavor-changing interactions and how precision lifetime measurements, which are the subject of this dissertation, help us understand them. In doing so we touch upon several Standard Model predictions as well as the theoretical tools that stand behind them. We conclude this chapter by laying out the major concepts, facts, and assumptions on which the analyses documented in this dissertation are based.

1.1 Weak interaction and the CKM description of the flavor sector

Originally the weak interaction was postulated to account for nuclear β decay. Later on, when pions were discovered to decay to muons, and muons to electrons, with characteristic times much too long to be attributed to strong or electromagnetic interactions, the weak interaction was invoked again. Eventually, observations in the weak decays of hadrons helped establish and advance the quark model. In the current section we give a short historical overview of this evolution, which is followed by a more formal discussion of the CKM picture of the quark sector and weak interactions.

1.1.1 Historical development

The Fermi theory of β decay [1], introduced in 1934, remained relatively solid for many years until parity non-conservation, which this theory did not describe, was pointed out by Lee and Yang [2] in 1956, and then experimentally proven by Wu and collaborators [3] in 1957. Sakurai [4] introduced a $V - A$ current structure to accommodate the observed parity violation, but the theory still suffered from the problem of non-renormalizability. In 1960s, mostly due to Glashow, Weinberg, and Salam [5–8] a new theory of electroweak interaction emerged. The new theory was based on the SU(2)×U(1) group and hypothesized four intermediate gauge fields: $\overrightarrow{W}_\mu$ and B_μ. These fields mix and acquire masses via Higgs mechanism[1] [9, 10], resulting in three massive physical vector bosons, $W^\pm$ and Z^0, and the massless γ.

Though the weak Lagrangian was originally formulated only for leptons, its development went hand-in-hand with the development of quark theory. In 1964 Gell-Mann [11] and Zweig [12] proposed that the observed hierarchy of hadrons is due to their compositness of smaller objects, quarks. They proposed three flavors of quarks: u, d, and s. The u quark would have electric charge $+2/3$, while d and s would have $-1/3$ each. The resulting quark composition for some of the most studied hadrons at the time is given in Table 1.1.

p	n	π^+	π^-	K^0	$\overline{K^0}$
uud	udd	$u\bar{d}$	$\bar{u}d$	$\bar{s}d$	$s\bar{d}$

Table 1.1: Quark composition of some of the hadrons when the quark model was first introduced.

A year earlier Cabibbo [13] analyzed weak hadronic currents, in particular those in which[2] $\Delta S = 0$, corresponding to decays such as $n \to pe\bar{\nu}_e$, and $\Delta S = 1$, corresponding to decays such as $K^+ \to \mu^+\nu_\mu$. He came to the conclusion that experimental data are much better described if one "rotates" the two currents away from each other by an angle $\theta_c = 0.257$.[3] Gell-Mann re-interpreted Cabibbo findings into the language of (not yet experimentally confirmed) quarks, proposing that the weak interaction couples the u quark to a mixed state $|d'\rangle = \cos\theta_c|d\rangle + \sin\theta_c|s\rangle$. With this definition all known weak interactions of the times could be described by a single coupling constant, which would be multiplied by $\sin\theta_c$ for $s \to u$ transitions and by $\cos\theta_c$ for $d \to u$ transitions.

Next Glashow, Iliopoulos, and Maiani [14] hypothesized the existence of the fourth quark, c, in order to explain the fact that the $K^0 \to \mu^+\mu^-$ decay rate was much smaller

[1] Higgs mechanism provides masses by means of interaction with the scalar Higgs field with non-zero vacuum expectation value, thus eliminating the need for explicit mass terms in the Lagrangian, which cause renormalizability problems. Unwanted Goldstone bosons are also avoided in the process.

[2] In Cabibbo's notation S was the "strangeness" quantum number, which is now simply identified with the number of s quarks.

[3] The value of θ_c was determined from the experimentally measured rates for $K^+ \to \mu^+\nu_\mu$ and $\pi^+ \to \mu^+\nu_\mu$ via relation $\Gamma(K^+ \to \mu^+\nu_\mu)/\Gamma(\pi^+ \to \mu^+\nu_\mu) = \tan^2\theta_c M_K(1 - M_\mu^2/M_K^2)^2/M_\pi(1 - M_\mu^2/M_\pi^2)^2$.

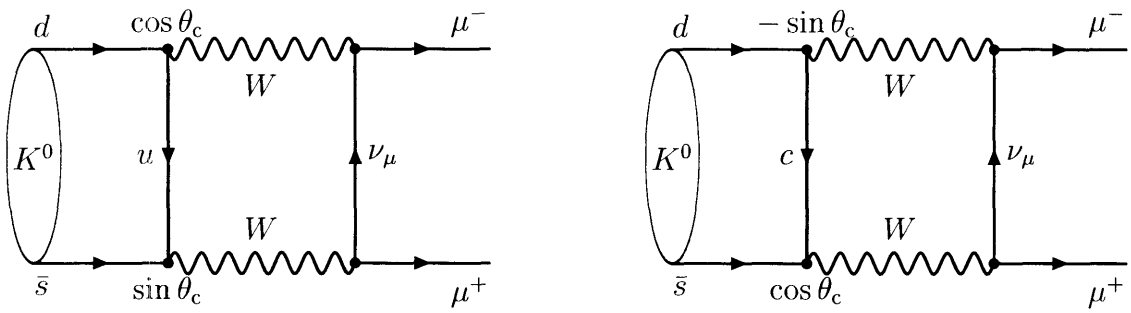


Figure 1-1: (a) The only first-order diagram for the $K^0 \to \mu^+\mu^-$ process before the GIM mechanism. (b) The other first-order diagram for the $K^0 \to \mu^+\mu^-$ process, the one involving the c quark, which nearly cancels the first one and makes the decay rate predicted by theory agree with experimental observations. In both cases factors multiplying the coupling constant are indicated for the $q_1\bar{q}_2W$ vertices.

than predicted by then-standard theory. Before them, there was only one first-order diagram, the one shown in Figure 1-1 (a), accounting for this decay. They proposed that a new quark, c, also couples only to a mixed state of s and d, though a different one. To generalize the concept of mixed quark states they introduced a mixing matrix

$$\begin{bmatrix} d' \\ s' \end{bmatrix} = \begin{bmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{bmatrix} \begin{bmatrix} d \\ s \end{bmatrix}, \tag{1.1}$$

which parametrizes the rotation of the (d, s) basis into (d', s'), with the latter being composed of the eigenstates of the weak interaction. Now there was an additional diagram contributing to the $K^0 \to \mu^+\mu^-$ process, the one shown in Figure 1-1 (b), which would perfectly cancel the first one if the masses of the c and u are equal. Because there is a difference in mass there is not a perfect cancellation, but still a suppression, which is known as GIM mechanism. By adjusting the c quark mass the predicted rate was made to match experimental observations.

It should be noted that in 1970, when Glashow, Iliopoulos, and Maiani came up with the GIM mechanism, the experimental indications of the existence of quarks, such as proton sub-structure [15] discovered in 1968, had already been on the scene. However, the quark model was still suspect, and the trio's proposal was not given a lot of attention.

Even less notice was given to the 1973 paper[4] by Kobayashi and Maskawa [16], who proposed extending the quark model from four quarks to six, *i.e.* from two generations to three, by adding the t and b quarks. Their primary motivation was to account for the still mysterious CP violation, observed at the 5σ level in the decays

[4]This paper is among the top three most cited papers of all times in the SPIRES-HEP database, as of the survey of 12/31/2002

of K^0 mesons in 1964 [17].

With three generations Kobayashi and Maskawa could incorporate CP violation by means of a mixing matrix for the down-like quarks (d, s, b)[5], which would have an analogous meaning to the matrix in Equation 1.1, namely, weak interaction couples a u quark to a linear combination of d, s, and b; the same for c and t, albeit the corresponding linear combinations of the down-like quarks are different ones. They demonstrated that unlike the case of four quarks which has a 2×2 matrix, any choice of the phase convention for the quark fields retains one complex phase in the unitary 3×3 mixing matrix for the case of six quarks. The presence of such a phase means that the Hamiltonian of the theory necessarily contains complex terms, which in turn means that such Hamiltonian is not invariant under T or, equivalently, CP symmetry.

Experimentalists had work to do, but they caught up quickly. In 1974 two different collaborations [18, 19] independently found a new particle, the J/ψ, which was soon identified as a $c\bar{c}$ meson. Then, in 1977, a $b\bar{b}$ bound state was discovered [20], and the massive intermediate vector bosons $W^{\pm}$ and Z^0 were observed [21] in 1983, making the discovery of the t quark just a matter of time. Time ran out in 1995, when the t quark was discovered [22, 23] and the modern picture of quarks and the weak interaction (except for the elusive Higgs boson(s)) had been settled.

1.1.2 Weak Lagrangian: the origin of the CKM Matrix

In this section we discuss in a more mathematical fashion how the CKM Matrix arises and what the consequences are.

In the Standard Model (SM) quark masses are generated via quark-Higgs interactions. The corresponding part of the SM Lagrangian for the case of the u' and d' quarks can be written (following the discussion in Reference [24]) as

$$\Delta\mathcal{L}_S^{quarks} = -\lambda_u \overline{\psi}'_L K u'_R - \lambda_d \overline{\psi}'_L K^{c.c.} d'_R + h.c., \tag{1.2}$$

where $\lambda_{u,d}$ are new dimensionless coupling constants, $\overline{\psi}'_L = [\bar{u}', \bar{d}']_L$ is the pair of left-handed Dirac-conjugated quark fields, and $K^{(c.c.)}$ is the (charge-conjugated) scalar (hence $\mathcal{L}_S$) Higgs doublet:

$$K = \frac{1}{\sqrt{2}} \begin{bmatrix} \chi \\ \sqrt{2} i\phi^- \end{bmatrix}, \qquad K^{c.c.} = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2} i\phi^+ \\ \chi^* \end{bmatrix}. \tag{1.3}$$

In the above and throughout this section we use primed quark notation for the *weak eigenstates* of the quarks, those which give the electro-weak interaction Lagrangian

[5] In reality, rotating the down-like quarks and leaving up-like quarks intact, is a pure convention. The approach of rotating up-like quarks, or both by "a little bit" has equal validity. This will become clear through the rest of the chapter.

its standard form:

$$\Delta\mathcal{L}_V^{quarks,i} = \sum_{q(uarks)} \left[ig \sin\theta_w Q_q A_\mu \bar{q}' \gamma_\mu q' + i\frac{g}{2\cos\theta_w} Z_\mu \bar{q}' \gamma_\mu (v_q + a_q\gamma_5) q' \right]$$
$$+ \sum_{g(enerations)} \left[i\frac{g}{2\sqrt{2}} W_\mu^+ \bar{u}_g' \gamma_\mu (1+\gamma_5) d_g' + i\frac{g}{2\sqrt{2}} W_\mu^- \bar{d}_g' \gamma_\mu (1+\gamma_5) u_g' \right]. \quad (1.4)$$

How the weak eigenstates differ from the *mass/flavor eigenstates* is in essence the subject of this section. The distinction will become clear toward its end.

Equation 1.2 can be rewritten as

$$\Delta\mathcal{L}_S^{quarks} = -\frac{g}{\sqrt{2}m_W} \overline{\psi}_L' m_u K u_R' + \frac{g}{\sqrt{2}m_W} \overline{\psi}_L' m_d K^{c.c.} d_R' + h.c., \quad (1.5)$$

in which we have established $\lambda_u = \frac{g}{\sqrt{2}}\frac{m_u}{m_W}$ and $\lambda_d = -\frac{g}{\sqrt{2}}\frac{m_d}{m_W}$ via identification of the $u\bar{u}$ and $d\bar{d}$ mass terms, m_W is the bare W mass, and g is the weak coupling constant. Equation 1.5 is easily generalized to the case of three quark generations, including inter-generational mixing:

$$\Delta\mathcal{L}_S^{quarks} = -\frac{g}{\sqrt{2}m_W} \overline{\Psi}_L' \mathbf{M}_U' K U_R' + \frac{g}{\sqrt{2}m_W} \overline{\Psi}_L' \mathbf{M}_D' K^{c.c.} D_R' + h.c., \quad (1.6)$$

where U_R' and D_R' are now columns containing right-handed up- and down-like quark fields:

$$U_R' = \begin{bmatrix} u' \\ c' \\ t' \end{bmatrix}_R, \qquad D_R' = \begin{bmatrix} d' \\ s' \\ b' \end{bmatrix}_R, \quad (1.7)$$

$\overline{\Psi}_L'$ is the string of left-handed Dirac-conjugated fields:

$$\overline{\Psi}_L' = [\bar{u}', \bar{c}', \bar{t}', \bar{d}', \bar{s}', \bar{b}']_L = [\overline{U}_L', \overline{D}_L'], \quad (1.8)$$

and $\mathbf{M}_{U,D}'$ are complex 3×3 matrices, of which we have no *a priori* knowledge. It can be shown that the Lagrangian in Equation 1.6 is also gauge invariant as is the single-generation Lagrangian in Equation 1.2.

Substituting the explicit Higgs field (Equations 1.3) into Equation 1.6 we obtain the generalized mass term (the interaction term is omitted):

$$\Delta\mathcal{L}_S^{quarks,m} = -\frac{g}{\sqrt{2}m_W} \overline{U}_L' \mathbf{M}_U U_R' - \frac{g}{\sqrt{2}m_W} \overline{D}_L' \mathbf{M}_D D_R' + h.c., \quad (1.9)$$

in which $\mathbf{M}_{U,D}$ are just their primed counterparts rescaled by the Higgs field vacuum expectation value.

In general, the matrices $\mathbf{M}_{U,D}$ are not diagonal. One can diagonalize them and achieve the standard form for the mass term of the Lagrangian by means of bi-unitary transformations:

$$\mathbf{M}_U = \mathbf{U}_L^\dagger \mathbf{m}_U \mathbf{U}_R, \qquad \mathbf{M}_D = \mathbf{D}_L^\dagger \mathbf{m}_D \mathbf{D}_R, \quad (1.10)$$

where $\mathbf{m}_U$ and $\mathbf{m}_D$ are diagonal 3×3 matrices and the $\mathbf{U}$ and $\mathbf{D}$ matrices involved are unitary. Plugging Equations 1.10 into Equation 1.9 we can identify the *mass eigenstates*:

$$U_L = \mathbf{U}_L U'_L, \qquad U_R = \mathbf{U}_R U'_R, \qquad D_L = \mathbf{D}_L D'_L, \qquad D_R = \mathbf{D}_R D'_R. \tag{1.11}$$

In terms of mass eigenstates Equation 1.4 now reads:

$$\begin{aligned}\Delta\mathcal{L}_V^{quarks,i} = &\sum_{q(uarks)} \left[ig \sin\theta_w Q_q A_\mu \bar{q}\gamma_\mu q + i\frac{g}{2\cos\theta_w} Z_\mu \bar{q}\gamma_\mu(v_q + a_q\gamma_5)q\right] \\ &+i\frac{g}{2\sqrt{2}} W_\mu^+ \overline{U}\gamma_\mu(1+\gamma_5)\mathbf{V}D + i\frac{g}{2\sqrt{2}} W_\mu^- \overline{D}\gamma_\mu(1+\gamma_5)\mathbf{V}^\dagger U,\end{aligned} \tag{1.12}$$

where $\mathbf{V} \equiv \mathbf{U}_L\mathbf{D}_L^\dagger$ and is, by convention, acting on the down-like quarks D. The unitary matrix $\mathbf{V}$, which comes about in this way, is commonly referred to as the CKM Matrix. The CKM Matrix arises naturally in the SM, simply as a result of the most general Yukawa interaction compatible with gauge invariance[6].

1.1.3 CKM Matrix and the Unitarity Triangle

In the previous section we have established that the CKM Matrix arises out of our desire to have the most general among reasonably simple theories, and therefore exists on equal footing with everything else in the SM Lagrangian. Below we discuss how many additional parameters the existence of the CKM Matrix brings into the theory.

In the most general form we can write

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \mathbf{V} \times \begin{bmatrix} d \\ s \\ b \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \times \begin{bmatrix} d \\ s \\ b \end{bmatrix}. \tag{1.13}$$

A complex $n \times n$ matrix has $2n^2$ real parameters. However, we already know that $\mathbf{V}$ is unitary, *i.e.*

$$\mathbf{V}^\dagger = \mathbf{V}^{-1}. \tag{1.14}$$

From Equation 1.14 it follows that

$$\mathbf{V}^\dagger\mathbf{V} = 1, \tag{1.15}$$

which amounts to n real (diagonal) and $n(n-1)/2$ complex (above *or* below diagonal) conditions, and therefore a total of only $2n^2 - 2\frac{n(n-1)}{2} - n = n^2$ free real parameters. Below we have spelled out *all* of the equations that arise from the matrix Equation 1.15

[6]It is also very instructive to notice that the absence of flavor-changing neutral currents in Equation 1.12 is just a consequence of the unitarity of matrices $\mathbf{U}$ and $\mathbf{D}$

applied in the three-dimensional case of Equation 1.13:

$$(\mathbf{V}^\dagger\mathbf{V})_{11}: \quad V_{ud}^* V_{ud} + V_{cd}^* V_{cd} + V_{td}^* V_{td} = 1, \tag{1.16}$$

$$(\mathbf{V}^\dagger\mathbf{V})_{22}: \quad V_{us}^* V_{us} + V_{cs}^* V_{cs} + V_{ts}^* V_{ts} = 1, \tag{1.17}$$

$$(\mathbf{V}^\dagger\mathbf{V})_{33}: \quad V_{ub}^* V_{ub} + V_{cb}^* V_{cb} + V_{tb}^* V_{tb} = 1, \tag{1.18}$$

$$(\mathbf{V}^\dagger\mathbf{V})_{12}: \quad V_{ud}^* V_{us} + V_{cd}^* V_{cs} + V_{td}^* V_{ts} = 0,$$

$$(\mathbf{V}^\dagger\mathbf{V})_{13}: \quad V_{ud}^* V_{ub} + V_{cd}^* V_{cb} + V_{td}^* V_{tb} = 0,$$

$$(\mathbf{V}^\dagger\mathbf{V})_{23}: \quad V_{us}^* V_{ub} + V_{cs}^* V_{cb} + V_{ts}^* V_{tb} = 0,$$

$$(\mathbf{V}^\dagger\mathbf{V})_{21}: \quad V_{us}^* V_{ud} + V_{cs}^* V_{cd} + V_{ts}^* V_{td} = 0, \tag{1.19}$$

$$(\mathbf{V}^\dagger\mathbf{V})_{31}: \quad V_{ub}^* V_{ud} + V_{cb}^* V_{cd} + V_{tb}^* V_{td} = 0, \tag{1.20}$$

$$(\mathbf{V}^\dagger\mathbf{V})_{32}: \quad V_{ub}^* V_{us} + V_{cb}^* V_{cs} + V_{tb}^* V_{ts} = 0. \tag{1.21}$$

Note that among the six off-diagonal equations three are just complex-conjugated copies of the other three, which is a consequence of the fact that hermitian conjugation ($\dagger$) is a product of complex conjugation ($*$) and transposition (T). Above we have chosen to number only the independent equations.

We can also independently shift the phase of each of the $2n$ quark fields without any observable being affected. In this way we can absorb $2n$ parameters describing $\mathbf{V}$ into a redefinition of the quark fields. The only caveat here is that shifting all quark field phases by the same amount leaves the matrix invariant, and therefore the actual number of observable parameters describing $\mathbf{V}$ is reduced by $2n-1$, not by $2n$, and is equal to $n^2 - (2n-1) = (n-1)^2$.

In the case of only two doublets of quarks, $\mathbf{V}$ is parametrized by a single real number, θ_c (see, for example, Equation 1.1). In the actual case of three quark generations there are four observable parameters that define the CKM Matrix. Had $\mathbf{V}$ been purely real, it would have been a 3×3 orthogonal matrix ($\mathbf{O}^T\mathbf{O} = 1$), describing a rotation in the three-dimensional space. Any such rotation is parametrized by three real parameters, *e.g.* Euler angles, and therefore the fourth parameter in $\mathbf{V}$ is necessarily an imaginary one, as intended by Kobayashi and Maskawa when they introduced the matrix (Section 1.1.1).

The following is considered the "standard" parameterization of the CKM matrix.

$$\mathbf{V} = \begin{bmatrix} c_{12}c_{13} & c_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{bmatrix}, \tag{1.22}$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$. Three angles θ_{12}, θ_{23}, and θ_{13} and the phase δ are the four parameters. These angles are defined in such a way that if one of them vanishes, so does the mixing between the respective two generations.

Reference [25] provides a number of arguments in favor of the parameterization in Equation 1.22. However, there is a popular approximation due to Wolfenstein [26], which emphasizes the hierarchy among the elements of the CKM Matrix, and is quite

precise despite being fairly simple:

$$\mathbf{V} = \begin{bmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ -iA^2\lambda^5\eta & 0 & 0 \\ A\lambda^5(\rho+i\eta)/2 & A\lambda^4(1/2-\rho-i\eta) & 0 \end{bmatrix} + \mathcal{O}(\lambda^6). \quad (1.23)$$

In this (approximate) parameterization $s_{12} = \lambda$, $s_{23} = A\lambda^2$, and $s_{13}e^{-i\delta} = A\lambda^3(\rho - i\eta)$. Here λ is small (≈ 0.22), while the other parameters are of the order of unity. The second term in Equation 1.23 is irrelevant for most practical purposes and is given only to illustrate the precision of the approximation.

Now let us go back to Equations 1.16–1.21 and notice that those that come from the off-diagonal elements of the product matrix $\mathbf{V}^\dagger\mathbf{V}$ represent triangles in the complex plane. Using Equation 1.23 it is straightforward to see that two of the triangles are extremely squashed, *i.e.* one side is much smaller than the other two, while the one defined by Equation 1.20 has all sides of the order of λ^3 and all angles reasonably large.

An important aside is that if we were to choose to translate the unitarity of $\mathbf{V}$ into $\mathbf{V}\mathbf{V}^\dagger = 1$ (and *not* $\mathbf{V}^\dagger\mathbf{V} = 1$ as we did in Equation 1.15) we would get a different set of equations in place of Equations 1.16–1.21 (though the first three would still be the same), which would also result in three triangles. However, these three triangles would generally be *different* from those we obtain with Equations 1.16–1.21, but there would still be one, corresponding to

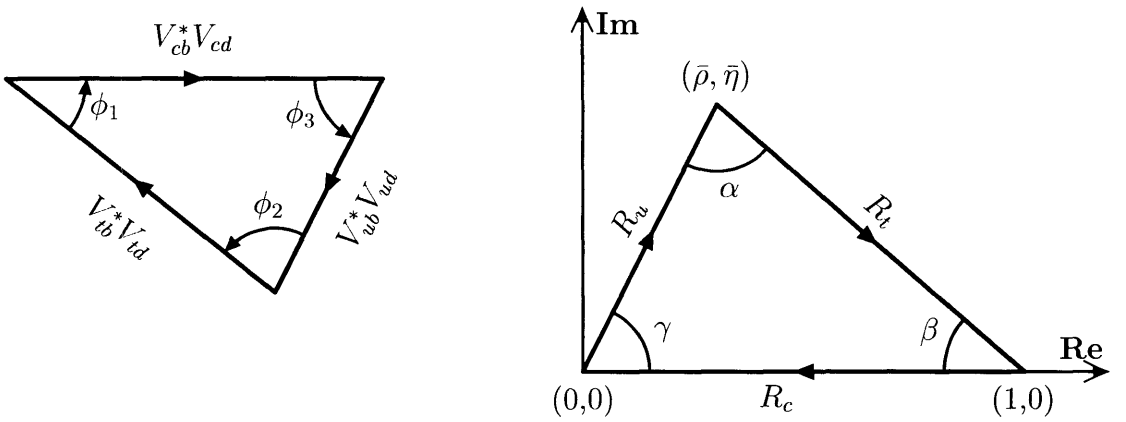
$$V_{ud}V_{td}^* + V_{us}V_{ts}^* + V_{ub}V_{tb}^* = 0, \quad (1.24)$$

with all sides of the order of λ^3.

At this time one of the priorities of the experimental particle physics is to "over-constrain" one such triangle, *i.e.* to measure all of its sides and angles and see if these make sense together. If they do, then the assumption of three generations and the unitarity of the CKM Matrix holds. If they do not, then physics beyond the Standard Model would be indicated.

From the experimental point of view triangles with large (consequently easier to measure) angles are preferred. Of the two such triangles described above, the one defined by Equation 1.20 is chosen. It is called "The Unitarity Triangle" and is shown in Figure 1-2 (a).

In practice it is more convenient to work with the normalized Unitarity Triangle. The rescaling is achieved when Equation 1.20 is divided by $V_{cb}^*V_{cd}$, so that one side of the Triangle becomes of unit length and is aligned with the real axis as shown in Figure 1-2 (b). Then the coordinates of the vertices of the Unitary Triangle become



(a) (b)

Figure 1-2: (a) The Unitarity Triangle which arises when the complex numbers in Equation 1.20 are represented by vectors in a complex plane. (b) The normalized Unitarity Triangle is obtained by dividing Equation 1.20 by $V_{cb}^* V_{cd}$. Labels are explained in the text.

$(0,0)$, $(1,0)$ and $(\bar{\rho}, \bar{\eta})$, where

$$\bar{\rho} = \Big(1 - \frac{\lambda^2}{2}\Big)\rho, \qquad \bar{\eta} = \Big(1 - \frac{\lambda^2}{2}\Big)\eta. \tag{1.25}$$

The three angles of the Unitarity Triangle are given by

$$\phi_1 \equiv \beta \equiv arg\Big(\frac{V_{cb}^* V_{cd}}{-V_{tb}^* V_{td}}\Big), \tag{1.26}$$

$$\phi_2 \equiv \alpha \equiv arg\Big(\frac{V_{tb}^* V_{td}}{-V_{ub}^* V_{ud}}\Big), \tag{1.27}$$

$$\phi_3 \equiv \gamma \equiv arg\Big(\frac{V_{ub}^* V_{ud}}{-V_{cb}^* V_{cd}}\Big) = \delta, \tag{1.28}$$

where $\{\phi_1, \phi_2, \phi_3\}$ and $\{\alpha, \beta, \gamma\}$ are the two naming conventions with the first being less popular, but less confusing at the same time.

The lengths of the sides of the normalized Unitary Triangle are given by

$$R_c \equiv 1, \tag{1.29}$$

$$R_u \equiv \left|\frac{V_{ub}^* V_{ud}}{V_{cb}^* V_{cd}}\right| = \sqrt{\bar{\rho}^2 + \bar{\eta}^2} = \Big(1 - \frac{\lambda^2}{2}\Big)\frac{1}{\lambda}\left|\frac{V_{ub}}{V_{cb}}\right|, \tag{1.30}$$

$$R_t \equiv \left|\frac{V_{tb}^* V_{td}}{V_{cb}^* V_{cd}}\right| = \sqrt{(1-\bar{\rho})^2 + \bar{\eta}^2} = \frac{1}{\lambda}\left|\frac{V_{td}}{V_{ts}}\right|. \tag{1.31}$$

It should be noted that the equality sign, "=", in Equations 1.25 – 1.31 implies

"equals up to the terms of $\mathcal{O}(\lambda^4)$."

Finally, the magnitude of the CP violating effects in the Standard Model is proportional [27] to the quantity $J = \mathrm{Im}(V_{ij}V_{kl}V_{kj}^*V_{il}^*) \propto \det[\mathbf{M}_U, \mathbf{M}_D]$. In the standard parameterization, given by Equation 1.22, $J = s_{12}s_{13}s_{23}c_{12}c_{13}^2c_{23}\sin\delta$, which is just twice the area of the Unitarity Triangle.

1.2 Brief theoretical overview of B meson lifetimes

A few powerful ways to probe the CKM Matrix involve measuring the b quark lifetime.

In the Standard Model the b quark decays to a c or u quark and a virtual W boson, with couplings given by the corresponding CKM Matrix elements. The W can decay

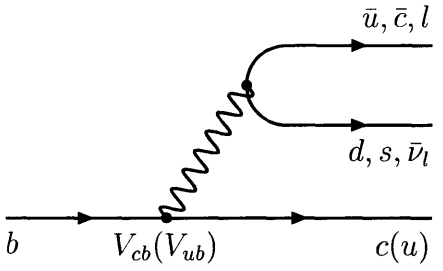


Figure 1-3: b quark decay in the Standard Model.

to $(\bar{u}, d)$ or $(\bar{c}, s)$ pair of quarks or a $(l, \bar{\nu}_l)$ lepton pair, as shown in Figure 1-3. The matrix element for the case of the leptonic W decay can be written as the product of quark and leptonic currents:

$$M_{l\bar{\nu}_l} \simeq -\frac{G_F}{\sqrt{2}} V_{qb} \cdot \bar{q}\gamma^\mu(1-\gamma_5)b \cdot \bar{l}\gamma_\mu(1-\gamma_5)\nu_l, \tag{1.32}$$

in which $G_F = \frac{g^2}{2\sqrt{2}m_W^2}$ is the Fermi coupling constant, and q is either c or u. The approximate equality in Equation 1.32 reflects the fact that we have replaced the true W propagator $\left(-\frac{1}{p^2-m_W^2}\right)$ with G_F — a perfectly valid thing to do, because $p < m_b \ll m_W$. Squaring the absolute value of $M_{l\bar{\nu}_l}$ and integrating over the phase space one obtains the leptonic decay width:

$$\Gamma_{l\bar{\nu}_l}(b \to q) = \frac{G_F^2 m_b^5}{192\pi^3} |V_{qb}|^2 F(\epsilon_q), \tag{1.33}$$

where $\epsilon_q = m_q/m_b$ and $F(\epsilon_q)$ is the phase space factor calculated [28] to be

$$F(\epsilon_q) = 1 - 8\epsilon_q^2 + \epsilon_q^6 - \epsilon_q^8 - 24\epsilon_q^4 \ln \epsilon_q. \tag{1.34}$$

Similar relationship is obtained for the case of hadronic W decay:

$$\Gamma_{q_1\bar{q}_2}(b \to q) = \frac{G_F^2 m_b^5}{192\pi^3} 3|V_{q_1q_2}|^2 |V_{qb}|^2 F(\epsilon_q), \tag{1.35}$$

in which $V_{q_1 q_2}$ is either V_{ud} or V_{cs}, depending on which way the W decays, and a factor of 3 accounts for the three possible color–anti-color combinations the $(q_1, \bar{q}_2)$ pair can possess. The gluon radiation corrections are not included. It is no coincidence that Equations 1.33 and 1.35 are reminiscent of the muon lifetime formula, because the calculation is essentially the same.

The total width of the b quark, Γ_b, is related to the partial width of a specific decay channel, $\Gamma_{chan.}$, via the branching fraction for this channel, $Br_{chan.}$:

$$\frac{Br_{chan.}}{\Gamma_{chan.}} = \frac{1}{\Gamma_b} \equiv \tau_b. \tag{1.36}$$

Thus, naively, one could pick out an exclusive or inclusive channel related to V_{ub} or V_{cb} and determine its branching fraction. Then, if τ_b is measured, the appropriate equation of the kind of Equations 1.33 and 1.35 can be used to extract the magnitude of the CKM Matrix element of interest. To measure the ratio $|V_{ub}/V_{cb}|$ entering Equation 1.30 one does not even need to know what τ_b is.

However, b quarks are only available for experimental study as constituents of the B hadrons. A great deal of additional effort is needed to relate the B hadron observables to the basic theory we have outlined. As we shall see, QCD effects arising in the presence of a lighter quark make precise quantitative predictions difficult. In particular, non-perturbative calculation techniques need to be invoked.

In what follows we discuss the lifetimes of the various B hadrons, the predictions for the differences among them as well as the theoretical instruments that are used to make these predictions.

1.2.1 B lifetimes: the spectator model and beyond

In the spectator model, a heavy quark Q (c or b) in a hadron H_Q is bound to the lighter "spectator" quark(s). For as long as the spectator ansatz holds, the decay of H_Q is governed by the weak decay of the Q, and, for this very reason, the lifetimes of all hadrons containing Q are the same and equal to that of a free Q. However, lifetimes of B hadrons have been experimentally observed to follow the pattern

$$\tau_{B_c} < \tau_{baryon} < \tau_{B_s} \simeq \tau_{B_d} < \tau_{B_u}. \tag{1.37}$$

On the phenomenological level, the main mechanisms generating the observed hierarchy of B hadron lifetimes are Pauli Interference, Weak Annihilation, and Weak Exchange.

To understand Pauli Interference (PI), consider diagrams in Figures 1-4 (a)–(d). For B_u decay, the external W emission, shown in Figure 1-4 (a), and the internal (color-suppressed) W emission, shown in Figure 1-4 (b), result in the same final state, therefore there is an interference between the two diagrams. The interference turns out to be destructive, explaining why the lifetime of B_u is longer than that of B_d, for which interference is absent because the corresponding final states are different, as shown in Figures 1-4 (c)–(d).

Weak Annihilation (WA) works to shorten the lifetimes of charged B mesons (B_u,

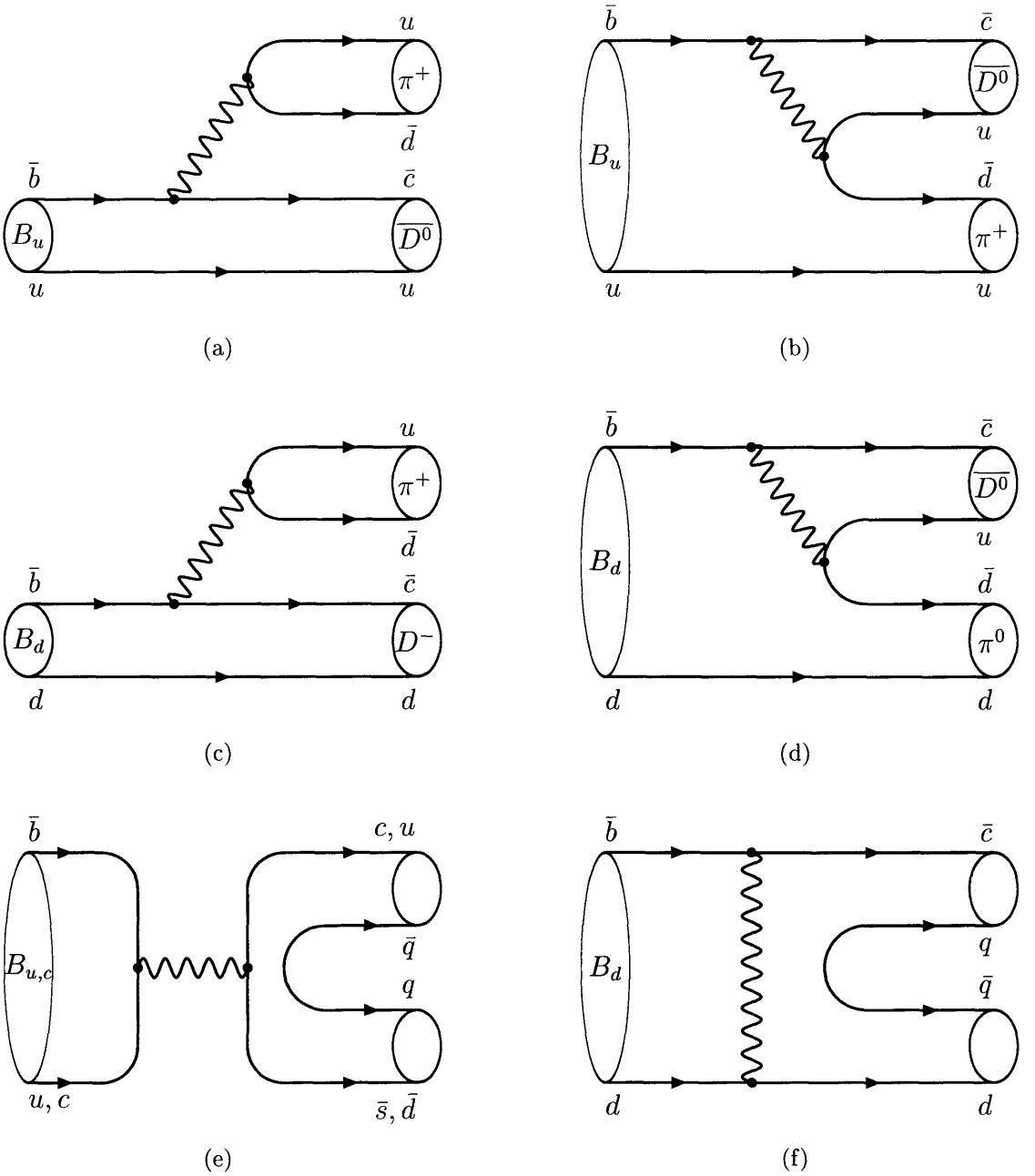


Figure 1-4: Lowest order diagrams showing decays of B flavored mesons.

B_c) by providing and additional mechanism for their decay, as shown in Figure 1-4 (e). Because there are no flavor changing neutral currents, the lifetimes of the B_d and B_s are unaffected by WA.

Weak Exchange (WE) diagrams, such as the one in Figure 1-4 (f), contribute to the lifetimes of B baryons and neutral B mesons. In the mesons such a contribution is helicity suppressed – the spins of the resulting quarks, $\bar{c}$ and d in the case of Figure 1-4 (f), must add up to zero (the spin of B_d), which means that both the

quark and the anti-quark have the same helicity. Such configuration is disfavored by the weak interaction, with, perhaps, a more familiar example being the prevalence of $\pi^- \to \mu^- \bar{\nu}_\mu$ decay over $\pi^- \to e^- \bar{\nu}_e$. For baryons such suppression is absent because there is no spin requirement on the two quarks, which is why WE is believed to be the explanation for B baryon lifetimes being slightly shorter than those of neutral B mesons.

1.2.2 HQE review

In this section we review the Heavy Quark Expansion (HQE). HQE is a theoretical framework that allows one to calculate total decay widths of heavy-flavored hadrons systematically including phenomena like those discussed in the previous sub-section. In our brief review we follow Reference [29]. A somewhat more elaborate discussion can be found in References [30, 31].

The heavy quark Q in a hadron H_Q is surrounded by a cloud of quarks, anti-quarks and gluons, with which Q itself and its decay products interact strongly. Until recently it had been a nearly impossible task to calculate how such an environment affects the lifetime of the Q. However, HQE provides a framework in which the calculations can be carried out based on first principles of QCD.

Starting from the optical theorem, the transition rate to the inclusive final state f can be written in terms of the imaginary part of the forward scattering operator:

$$\hat{T}(Q \to f \to Q) = i\mathrm{Im} \int T\{\mathcal{L}_W(x)\mathcal{L}_W^\dagger(0)\}d^4x, \tag{1.38}$$

where $T\{\}$ stands for time-ordered product and $\mathcal{L}_W$ is the effective weak Lagrangian. Under certain conditions the non-local operator product in Equation 1.38 can be decomposed into an infinite sum of local operators O_i of increasing dimension i. The rate (width) of $H_Q \to f$ transition is then obtained by calculating the expectation value of $\hat{T}$ for the state H_Q:

$$\langle H_Q|\hat{T}(Q \to f \to Q)|H_Q\rangle \propto \Gamma(H_Q \to f) = |CKM|^2 \sum_i \tilde{c}_i^{(f)} \langle H_Q|O_i|H_Q\rangle, \tag{1.39}$$

where $|CKM|^2$ is the relevant combination of the CKM Matrix elements. The coefficients $\tilde{c}_i^{(f)}$ in Equation 1.39 are known as Wilson coefficients, after K. Wilson, who proposed [32] the Operator Product Expansion in 1969. They contain higher and higher powers of $\frac{1}{m_Q}$, which multiply correspondingly higher powers of Λ_{QCD} contained in the expectation values of operators O_i. The $\tilde{c}_i^{(f)}$'s can be computed in perturbation theory. Hence, the HQE predicts $\Gamma(H_Q \to f)$ in the form of an expansion in both Λ_{QCD}/m_Q and $\alpha_s(m_Q)$. The precision of current experiments makes it mandatory to go to the next-to-leading order in QCD, *i.e.* to include corrections of the order of $\alpha_s(m_Q)$ to the $\tilde{c}_i^{(f)}$'s, when HQE predictions are made.

All non-perturbative physics is shifted into the expectation values $\langle H_Q|O_i|H_Q\rangle$ of operators O_i. They can be calculated using lattice QCD or QCD sum rules. Al-

ternatively, they could be related to other observables via HQE, as shown in detail in Reference [29]. One may reasonably expect that powers of $\frac{\Lambda_{QCD}}{m_Q}$ provide enough suppression so that only the first few terms of the sum in Equation 1.39 matter.

For the case of semileptonic and hadronic decays Equation 1.39 can be expanded as follows

$$\begin{aligned}\Gamma(H_Q \to f) = \frac{G_F^2 m_Q^5}{192\pi^3}|CKM|^2 \Bigg[& c_3^{(f)} \langle H_Q|\overline{Q}Q|H_Q\rangle \\ & + c_5^{(f)} \frac{\langle H_Q|\overline{Q}i\sigma GQ|H_Q\rangle}{m_Q^2} \\ & + \sum_i c_{6,i}^{(f)} \frac{\langle H_Q|(\overline{Q}\Gamma_i q)(\bar{q}\Gamma_i Q)|H_Q\rangle}{m_Q^3} + \mathcal{O}\Big(\frac{\Lambda_{QCD}^4}{m_Q^4}\Big)\Bigg],\end{aligned} \tag{1.40}$$

where we have chosen to write Wilson coefficients' dependence on m_Q and the familiar $\frac{G_F^2 m_b^5}{192\pi^3}|CKM|^2$ factor explicitly. Using the equations of motion one finds that

$$\overline{Q}Q = \overline{Q}\gamma_0 Q - \frac{\overline{Q}[(iD)^2 - \frac{i}{2}\sigma G]Q}{2m_Q^2} + \cdots . \tag{1.41}$$

From Equation 1.41, taking into account that $\overline{Q}\gamma_0 Q$ is a conserved current, one immediately concludes that

$$\langle H_Q|\overline{Q}Q|H_Q\rangle = 1 + \mathcal{O}\Big(\frac{\Lambda_{QCD}^2}{m_Q^2}\Big). \tag{1.42}$$

Reference [29] goes on to evaluate other expectation values in terms of observables. For example, for the expectation value of the chromomagnetic operator, which enters the expansion in Equation 1.41, and on its own right in Equation 1.40, one obtains:

$$\langle P_Q|\overline{Q}i\sigma GQ|P_Q\rangle \simeq \frac{3}{2}\big(M_{V_Q}^2 - M_{P_Q}^2\big), \tag{1.43}$$

where P_Q stands for a pseudo-scalar meson (B), and V_Q stands for a vector meson (B^*). However, certain expectation values cannot be determined in such a way. For instance, only the difference between baryon and meson states can be reliably estimated for $\langle \boldsymbol{p}_Q^2\rangle_{H_Q} \equiv \langle H_Q|(iD)^2|H_Q\rangle$:

$$\langle \boldsymbol{p}_Q^2\rangle_{\Lambda_Q} - \langle \boldsymbol{p}_Q^2\rangle_{P_Q} \simeq \frac{2m_b m_c}{m_b - m_c}\Big[\big(\langle M_B\rangle - M_{\Lambda_b}\big) - \big(\langle M_D\rangle - M_{\Lambda_c}\big)\Big], \tag{1.44}$$

where $\langle M_{B(D)}\rangle \equiv \frac{1}{4}\big(M_{B(D)} + 3M_{B^*(D^*)}\big)$ is the "spin averaged" meson mass.

Present HQE results are of somewhat limited precision, which in some instances is already surpassed by the experimental measurements, *e.g.* in the case of the B_u and B_d lifetime ratio as shown in Table 1.2. More accurate predictions are a matter

of progress in the evaluation of the non-perturbative hadronic matrix elements and, not surprisingly, better experimental input. However, HQE even in its present shape draws a number of important conclusions, which are in agreement with experimental observations:

- The heavier the Q the smaller is the variation in lifetimes among different H_Q's, which is to say that as $m_Q \to \infty$ we retrieve the spectator picture in which the lifetimes of all Q-flavored hadrons are the same.
- The non-perturbative corrections only arise at the order of Λ^2_{QCD}/m^2_Q, which translates in 5-10% differences among B hadron lifetimes.
- Since light quark fields (q) only enter at the Λ^3_{QCD}/m^3_Q level the lifetimes of all the B mesons are practically the same up to Λ^3_{QCD}/m^3_Q corrections. It is largely only the difference between meson and baryon lifetimes that appears at the Λ^2_{QCD}/m^2_Q level, yet the Λ^3_{QCD}/m^3_Q corrections are enhanced by a phase space factor $16\pi^2$ with respect to the leading free Q decay.

Table 1.2, compiled from References [25,33–36], compares a number of HQE predictions with data. There is a quantitative agreement between the predictions and the measurements for $Q = b$ and even a semi-quantitative agreement for $Q = c$, where one would not necessarily expect it, because Λ_{QCD}/m_c is a notably worse expansion parameter than Λ_{QCD}/m_b.

Quantity	HQE Prediction	Data
τ_{D^+}/τ_{D^0}	~ 2	2.55 ± 0.034
$\tau_{D_s^+}/\tau_{D^0}$	1.08 ± 0.04	1.125 ± 0.042
$\tau_{\Lambda_c^+}/\tau_{D^0}$	~ 0.5	0.489 ± 0.008
$\tau_{\Xi_c^+}/\tau_{\Lambda_c^+}$	~ 1.3	1.75 ± 0.36
$\tau_{\Xi_c^+}/\tau_{\Xi_c^0}$	~ 2.8	3.57 ± 0.91
$\tau_{\Xi_c^+}/\tau_{\Omega_c^0}$	~ 4	3.9 ± 1.7
τ_{B_u}/τ_{B_d}	1.053 ± 0.023	1.074 ± 0.014
τ_{B_s}/τ_{B_d}	1.00 ± 0.01	0.948 ± 0.038
$\Delta\Gamma_s/\Gamma_s$	$(\frac{f_{B_s}}{245\,\mathrm{MeV}})^2(0.12 \pm 0.04)$	< 0.29 (95% C.L.)
$\tau_{\Lambda_b}/\tau_{B_d}$	0.90 ± 0.05	0.796 ± 0.054

Table 1.2: HQE predictions vs. data both in c and b sectors.

Another strong point of HQE is that it makes an excellent laboratory out of B hadrons in which QCD effects can be studied. To illustrate this, the HQE prediction for $\Delta\Gamma_s/\Gamma_s$ in Table 1.2 is expressed in terms of the B_s meson decay constant f_{B_s}, which enters many other QCD based calculations, in particular those of Section 1.3.2.

In this dissertation we make a number of measurements that are relevant to HQE. These include the measurements of the lifetimes of the B_u, B_d, and B_s mesons and the corresponding lifetime ratios. On the other hand we benefit from HQE in that one of the results for the B_s fractional width difference, $\Delta\Gamma_s/\Gamma_s$, makes use of the HQE prediction that B_s and B_d total decay widths are equal to within 1%.

1.3 Mixing phenomenon in neutral B mesons

As has been pointed out in the preceding sections, had the CKM Matrix been diagonal there would not have been any inter-generational transitions. For example, the t quark would have decayed only to b, and the latter would have been stable. Off-diagonal elements of the CKM matrix, though small, do couple different generations. Only due to them are the B mesons allowed to decay, still very slowly. Another interesting phenomenon that a non-diagonal CKM Matrix makes possible is the neutral meson mixing, which we describe in this section for the case of B mesons.

1.3.1 Quantum-mechanical treatment

Consider a neutral B meson state $|B_q^0\rangle$ $(\bar{b}q)$. Here $q = s, d$, but we omit the index because the general discussion we present in the current section applies equally to B_d and B_s mesons. In the absence of weak flavor-changing interactions the $|B^0\rangle$ state $(\bar{b}q)$ and $|\overline{B^0}\rangle$ state $(b\bar{q})$ would have been degenerate eigenstates of the Hamiltonian. With the inclusion of weak (flavor mixing!) terms, the Hamiltonian is no longer diagonal in the $\{|B^0\rangle, |\overline{B^0}\rangle\}$ basis, and transitions between the two states are allowed. This phenomenon is known as *B meson mixing.* The lowest order mixing diagrams are shown in Figure 1-5.

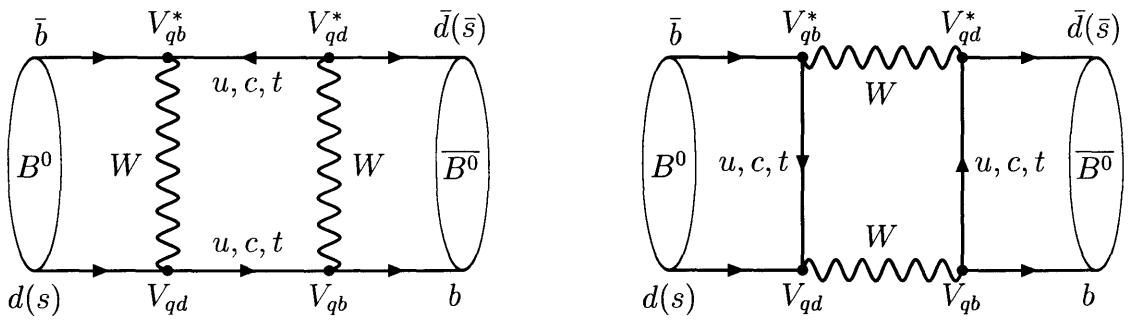


Figure 1-5: Lowest order Feynman diagrams, leading to $B_{d(s)}^0 - \overline{B^0}_{d(s)}$ meson mixing. These are also commonly known as *box diagrams.* Loop processes are dominated by t quark because their amplitudes are proportional to the square of the quark mass.

b quarks are typically produced by strong or electromagnetic interactions, therefore neutral B mesons come about in one of the two *flavor eigenstates*, $|B^0\rangle$ or $|\overline{B^0}\rangle$. Due to the mixing processes, after a time t the original pure state evolves into a quantum superposition of the two flavor eigenstates, $a(t)|B^0\rangle + b(t)|\overline{B^0}\rangle$.

In the Wigner-Weisskopf approximation [37] the evolution is governed by the Schrödinger equation:

$$i\frac{\partial}{\partial t}\begin{bmatrix} a(t) \\ b(t) \end{bmatrix} = \left(\mathbf{M} - \frac{i}{2}\mathbf{\Gamma}\right) \times \begin{bmatrix} a(t) \\ b(t) \end{bmatrix}, \tag{1.45}$$

where $\mathbf{M}$ and $\mathbf{\Gamma}$ are the 2×2 mass and decay matrices. These matrices are Hermitian. CPT invariance requires

$$M_{11} = M_{22} \equiv M, \qquad \Gamma_{11} = \Gamma_{22} \equiv \Gamma, \tag{1.46}$$

with index 1 indicating B^0 and 2 indicating $\overline{B^0}$. The elements of the matrices can be obtained by calculating the amplitudes encoded in the diagrams of Figure 1-5.

The most general form of the eigenstates of the effective Hamiltonian in Equation 1.45 is:

$$|B_{L,H}\rangle = p|B^0\rangle \pm q|\overline{B^0}\rangle, \qquad |p|^2 + |q|^2 = 1. \tag{1.47}$$

We call these the *mass eigenstates* and label them accordingly: L (light) and H (heavy). Solving the eigenvalue problem we obtain (using notations of Equation 1.46):

$$\lambda_{L,H} = (M - \frac{i}{2}\Gamma) \pm \frac{q}{p}(M_{12} - \frac{i}{2}\Gamma_{12}) \tag{1.48}$$

with

$$\frac{q}{p} = \sqrt{\frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}}}. \tag{1.49}$$

Integrating Equation 1.45, we see that the mass eigenstates evolve in time with exponentials:

$$|B_{L,H}(t)\rangle = |B_{L,H}(0)\rangle e^{-i\lambda_{L,H}t} = |B_{L,H}(0)\rangle e^{-iM_{L,H}t - \frac{1}{2}\Gamma_{L,H}t}, \tag{1.50}$$

where in the last step we have defined

$$M_{L,H} = \mathrm{Re}(\lambda_{L,H}), \qquad \Gamma_{L,H} = -2\mathrm{Im}(\lambda_{L,H}). \tag{1.51}$$

Solving Equations 1.47 for $|B^0\rangle$ and $|\overline{B^0}\rangle$ and using Equations 1.50 we can derive the time-evolution of the pure flavor eigenstates:

$$|B^0(t)\rangle = g_+(t)|B^0\rangle + \frac{q}{p}g_-(t)|\overline{B^0}\rangle, \tag{1.52}$$

$$|\overline{B^0}(t)\rangle = g_+(t)|\overline{B^0}\rangle + \frac{p}{q}g_-(t)|B^0\rangle, \tag{1.53}$$

in which

$$g_\pm(t) = \frac{1}{2}(e^{-i\lambda_L t} \pm e^{-i\lambda_H t}). \tag{1.54}$$

In Equations 1.52 and 1.53 we have denoted by $|\overline{B^0}(t)\rangle$ the state at time $t > 0$ that was created as $|\overline{B^0}\rangle$ at time $t = 0$.

We can now calculate the probabilities of observing (via decay) the initial states to change flavor (mix) or to stay unchanged at a time t:

$$|\langle B^0|\overline{B^0}(t)\rangle|^2 = \left|\frac{p}{q}\right|^2 |g_-(t)|^2, \tag{1.55}$$

$$|\langle\overline{B^0}|B^0(t)\rangle|^2 = \left|\frac{q}{p}\right|^2 |g_-(t)|^2, \tag{1.56}$$

$$|\langle B^0|B^0(t)\rangle|^2 = |g_+(t)|^2, \tag{1.57}$$

$$|\langle\overline{B^0}|\overline{B^0}(t)\rangle|^2 = |g_+(t)|^2, \tag{1.58}$$

where

$$|g_\pm(t)|^2 = \frac{1}{2}\Big[\cosh\Big(\frac{\Delta\Gamma t}{2}\Big) \pm \cos(\Delta \mathrm{M} t)\Big] e^{-\Gamma t}, \tag{1.59}$$

in which we have defined

$$\Delta \mathrm{M} = M_H - M_L, \qquad \Delta\Gamma = \Gamma_L - \Gamma_H. \tag{1.60}$$

Similarly, the time-dependent rate of $\overline{B^0}(t)$ decay to almost any specific final state f depends both on ΔM and $\Delta\Gamma$. In CKM Matrix related measurements knowledge of ΔM and $\Delta\Gamma$ is often required to relate the observed CP asymmetry

$$A_{CP}^{(f)}(t) = \frac{\Gamma(B^0(t) \to f) - \Gamma(\overline{B^0}(t) \to f)}{\Gamma(B^0(t) \to f) + \Gamma(\overline{B^0}(t) \to f)} \tag{1.61}$$

to the CP violating phase involved. ΔM and $\Delta\Gamma$ themselves involve combinations of the CKM Matrix elements and could be used to constrain the Unitarity Triangle.

In the following section we present the results of the Standard Model calculation of ΔM and $\Delta\Gamma$ both for B_d and B_s systems, some implications, and a discussion of the status of the current mixing measurements as well as potential outcomes of future measurements.

1.3.2 Phenomenology

The dispersive (mass) part of the box diagrams, shown in Figure 1-5, is dominated by the t quark contribution, whereas the absorptive (lifetime) part is determined by the real intermediate states, which correspond to the common decay products of the B^0 and $\overline{B^0}$. Calculation of M_{12} and Γ_{12}, made possible by the B^0 mass being significantly away from the region of hadronic resonances [25,38], yields the following results [39]:

$$M_{12} = -\frac{G_F^2 m_W^2 \eta_{B_q} M_{B_q} f_{B_q}^2 B_{B_q}}{12\pi^2} S_0(m_t^2/m_W^2)(V_{tq}^* V_{tb})^2, \tag{1.62}$$

$$\Gamma_{12} = \frac{G_F^2 m_b^2 \eta'_{B_q} M_{B_q} f_{B_q}^2 B_{B_q}}{8\pi}\Big[(V_{tq}^* V_{tb})^2 + V_{tq}^* V_{tb} V_{cq}^* V_{cb}\mathcal{O}\Big(\frac{m_c^2}{m_b^2}\Big) + (V_{cq}^* V_{cb})^2 \mathcal{O}\Big(\frac{m_c^4}{m_b^4}\Big)\Big], \tag{1.63}$$

where f_{B_q} and B_{B_q} are the weak decay constant and the bag parameter, respectively; $S_0(x)$ is a known function approximated rather well by $0.784\, x^{0.76}$; and η_{B_q} and η'_{B_q} are QCD corrections which are of the order of unity.

The above two equations cannot be used for reliable quantitative estimates because

of a large uncertainty in present calculations of $f_{B_q}^2 B_{B_q}$. However, in the ratio of the two quantities, Γ_{12}/M_{12}, the largest uncertainties cancel, and one can still reliably conclude that

$$\left|\frac{\Gamma_{12}}{M_{12}}\right| \simeq \frac{3\pi}{2}\frac{m_b^2}{m_W^2}\frac{1}{S_0(m_t^2/m_W^2)} \sim \mathcal{O}\Big(\frac{m_b^2}{m_t^2}\Big) \tag{1.64}$$

and

$$\phi_{12} \equiv arg\Big(-\frac{M_{12}}{\Gamma_{12}}\Big) \sim \mathcal{O}\Big(\frac{m_c^2}{m_b^2}\Big). \tag{1.65}$$

We can use Equation 1.64, showing the smallness of $\left|\frac{\Gamma_{12}}{M_{12}}\right|$, and Equation 1.65 to perform the power expansion of $\frac{q}{p}$ (Equation 1.49) and the eigenvalues (Equations 1.48) to obtain:

$$\frac{q}{p} = -e^{-i\phi_M}\Big[1 - \frac{1}{2}\left|\frac{\Gamma_{12}}{M_{12}}\right|\sin\phi_{12}\Big] + \mathcal{O}\Big(\left|\frac{\Gamma_{12}}{M_{12}}\right|^2\Big), \tag{1.66}$$

$$\Delta\mathrm{M} = 2|M_{12}|\Big[1 + \mathcal{O}\Big(\left|\frac{\Gamma_{12}}{M_{12}}\right|^2\Big)\Big], \tag{1.67}$$

$$\Delta\Gamma = 2|\Gamma_{12}|\cos\phi_{12}\Big[1 + \mathcal{O}\Big(\left|\frac{\Gamma_{12}}{M_{12}}\right|^2\Big)\Big], \tag{1.68}$$

where $\phi_M \equiv arg(M_{12})$.

There are a number of things to take a notice of at this point:

- $\left|\frac{q}{p}\right| = 1$ holds to a very good approximation. In fact, Reference [40] estimates $1 - \left|\frac{q}{p}\right|^2 \sim \mathcal{O}(10^{-3})$ for $B_d - \overline{B}_d$ system and $\sim \mathcal{O}(10^{-4})$ for $B_s - \overline{B}_s$ system. If $\Delta\Gamma$ can be ignored (usually a good approximation for B_d), Equations 1.55 – 1.59 simplify to:

$$|\langle B^0|\overline{B^0}(t)\rangle|^2 = |\langle\overline{B^0}|B^0(t)\rangle|^2 = \frac{1}{2}\Big[1 - \cos(\Delta\mathrm{M}t)\Big]e^{-\Gamma t}, \tag{1.69}$$

$$|\langle B^0|B^0(t)\rangle|^2 = |\langle\overline{B^0}|\overline{B^0}(t)\rangle|^2 = \frac{1}{2}\Big[1 + \cos(\Delta\mathrm{M}t)\Big]e^{-\Gamma t}. \tag{1.70}$$

- To an excellent precision

$$\Delta\mathrm{M} = 2|M_{12}|, \qquad \Delta\Gamma = 2|\Gamma_{12}|\cos\phi_{12} \simeq 2|\Gamma_{12}|, \tag{1.71}$$

  with the right-hand sides given by Equations 1.62 and 1.63. In the last of the above equations we have neglected the difference between $\cos\phi_{12}$ and unity which in the Standard Model is of the order of $m_c^4/m_b^4 \approx 0.007$. These equations make it explicit that both $\Delta\mathrm{M}$ and $\Delta\Gamma$ are non-negative quantities in the Standard Model, for the way we have defined them.

- Using Equations 1.62 and 1.66 in conjunction with the results of Section 1.1.3, we find that to a very good approximation:

  – For B_d, $\frac{q}{p} = -e^{-2i\beta}$.

– For B_s, $\frac{q}{p} = -1$, which means that mass eigenstates are nearly CP eigenstates [41, 42]:

$$CP|B_s^L\rangle = +|B_s^L\rangle, \qquad CP|B_s^H\rangle = -|B_s^H\rangle. \tag{1.72}$$

The last point is by far the most important implication of all, at least for the purposes of this dissertation.

In many cases it is convenient to deal with the dimensionless quantities $x \equiv \Delta M/\Gamma$ and $2y \equiv \Delta\Gamma/\Gamma$, where the average width, $\Gamma \equiv (\Gamma_L + \Gamma_H)/2$, is chosen as the normalization. In the $B_d - \overline{B}_d$ system the mass difference has been measured [25] quite precisely:

$$x_d = 0.772 \pm 0.013. \tag{1.73}$$

The fractional width difference is calculated [43] to be very small, $(3.0 \pm 1.2) \times 10^{-3}$, and a recent measurement [44] has obtained

$$\frac{\Delta\Gamma_d}{\Gamma_d} = -0.008 \pm 0.037(stat.) \pm 0.018(syst.), \tag{1.74}$$

compatible with the prediction.

The $B_s - \overline{B}_s$ system is to a large extent unexplored territory. Only bounds [25] exist from experimental studies:

$$x_s > 20.0 \quad (95\% \text{ C.L.}), \tag{1.75}$$

$$\frac{\Delta\Gamma_s}{\Gamma_s} < 0.29 \quad (95\% \text{ C.L.}). \tag{1.76}$$

With ΔM_d precisely known, the measurement of ΔM_s becomes very important. In the ratio of ΔM_d and ΔM_s, which can be constructed using Equations 1.71 and 1.62, the largest hadronic uncertainties almost cancel, allowing one to cleanly extract the magnitude of the ratio of CKM elements $|V_{td}/V_{ts}|$, thereby constraining one side of the unitarity triangle (Figure 1-2 (b), Equation 1.31).

The ΔM_s measurement can potentially be accomplished in a way similar to that of ΔM_d — by measuring the frequency of flavor oscillations in decays of B_s mesons governed by Equations 1.55 – 1.58. Such an approach requires excellent proper decay length resolution, as well as the ability to tag the flavor, *i.e.* distinguish between particle and anti-particle, of the B_s meson at decay and production times. The latter can be a rather inefficient process. Furthermore, it may turn out that oscillations are simply too fast, *i.e.* ΔM_s is too large, to be resolved by state-of-the-art detectors.

On the other hand, because of the proportionality between ΔM_s and $\Delta\Gamma_s$ stemming from Equations 1.62 and 1.63, the fractional width difference in $B_s - \overline{B}_s$ system is expected to be much larger than in $B_d - \overline{B}_d$ system. $\Delta\Gamma_s/\Gamma_s$ may turn out to be large enough to be measured with significant experimental precision. Current theory estimates [35, 45] prefer the range 0.12 ± 0.06. It is conceivable that Nature has made ΔM_s so large that for the time being $\Delta\Gamma_s/\Gamma_s$ is our only experimental insight into $B_s - \overline{B}_s$ mixing.

In another scenario, when both x_s and $\Delta\Gamma_s/\Gamma_s$ are measured with adequate precision, the ratio of the two can be a cross-check of theory. This becomes especially interesting in view of improving calculation techniques and shrinking theoretical uncertainties.

Yet another scenario for $\Delta\Gamma_s/\Gamma_s$ is when the upper bound is set as a result of the measurement, such that $\Delta\Gamma_s/\Gamma_s$ is incompatible with the prediction. As argued in References [30, 45], new physics would modify the calculation of the box diagrams, likely shifting ϕ_{12} away from zero. If $\Delta\Gamma_s/\Gamma_s$ is too small, in accordance with the second of Equations 1.71, it would necessarily mean large ϕ_{12} and would thus indicate physics beyond the Standard Model.

1.4 Angular correlations in the decays of neutral B mesons

With the current section we further the transition from fairly abstract theory to more practical matters of this dissertation. One such practical matter is how one separates the decays of the light $B_{d,s}$ mass eigenstates from those of the heavy ones. It turns out to be possible for the B_s!

Both neutral B decays considered in the present analysis, $B_d \to J/\psi K^{*0}$ and $B_s \to J/\psi\phi$, are decays of a pseudo-scalar to a vector$-$vector intermediate state. In such decays one can statistically distinguish those of different P parity by looking at the angular correlations among the final state particles. For $B_s \to J/\psi\phi$ however, since both J/ψ and ϕ are C-odd eigenstates, the properties of the $J/\psi\phi$ state under P are the same as those under CP. For this reason it makes sense to speak of separating $B_s \to J/\psi\phi$ decays of different CP. We have seen in Section 1.3.2 that in $B_s - \overline{B}_s$ system CP eigenstates nearly coincide with mass eigenstates, therefore angular analysis allows one, at least in principle, to separate decays of the light mass eigenstates from those of the heavy mass eigenstates. This opens up a possibility of measuring $\Delta\Gamma_s/\Gamma_s$. The current section describes a theoretical framework which provides a foundation for this measurement.

1.4.1 Transversity basis and transversity variables

One of the most suitable coordinate bases, both from theoretical and experimental points of view, to investigate the angular correlations among the final state particles in $P \to VV$ decays is the transversity basis. This is defined as follows.

Consider the decay chain $B_d \to J/\psi K^{*0} \to \mu^+\mu^- K^+\pi^-$. The transversity basis is fixed in the rest frame of the J/ψ. The K^{*0} flight direction defines the positive x axis, while the plane of $K^+\pi^-$ system defines the $x-y$ plane with the y axis oriented such that $p_y(K^+) > 0$. The two-fold ambiguity in choosing the z axis is solved by adopting a right-handed coordinate system. The angle θ is defined as the angle between μ^+ flight direction and the positive direction of the z axis as shown in Figure 1-6. Angle ϕ is the angle between the x-axis and the projection of the μ^+ direction onto the $x-y$ plane. Finally, the ψ angle is the angle of the K^+ in the K^{*0} rest frame relative

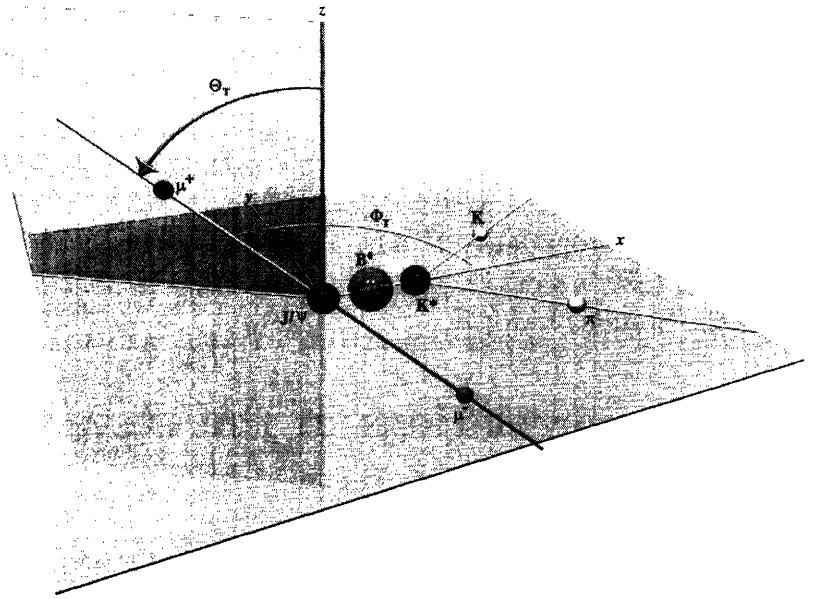


Figure 1-6: Transversity basis and angle definitions for the case of $B_d \to J/\psi K^{*0}$ decay. Angle ψ is not shown because it is defined in the different reference frame (see text). This definition translates trivially to the case of $B_s \to J/\psi\phi$ decay.

to the negative direction of the J/ψ in that frame. For the ease of writing formulae, we denote $\vec{\omega} = \{\cos\theta, \phi, \cos\psi\}$ the set of the three angular variables. This vector notation for $\vec{\omega}$ implies multiple components, but does not imply the transformation properties of a vector.

1.4.2 Time-dependent angular distributions in the $B_d \to J/\psi K^{*0}$ and $B_s \to J/\psi\phi$ decays

There are three decay amplitudes, corresponding to linear polarizations of the vector mesons (J/ψ, K^{*0}/ϕ), which are either longitudinal (A_0), or transverse to their flight direction and parallel ($A_{||}$) or perpendicular ($A_\perp$) to one another [46]. Throughout this dissertation we adopt Greek subscript α, which runs over the three values $\{0, ||, \perp\}$, such that whenever we need to refer to all three amplitudes we can write $\{A_\alpha\}$, while A_α denotes one/any of them.

Only the *relative* phases of the amplitudes can enter physics observables, so we are free to fix the phase of one of them. We adopt the convention in which $arg(A_0) = 0$ and define $\delta_{||} \equiv arg(A_{||})$, $\delta_\perp \equiv arg(A_\perp)$. These phases are expected to be close to 0 or π in the absence of significant final-state interactions [47].

The amplitude $A_\perp$ results in decays in which the two vector mesons are emitted with relative orbital angular momentum $L = 1$ and is thus associated with the P-odd decays. The amplitudes A_0 and $A_{||}$ are associated with (mixtures of) the $L = 0$ and

2 decays and are P-even [46]. It is assumed that the magnitudes of the corresponding decay amplitudes are equal for particle or anti-particle decay.

Reference [47] gives the differential rate at time t for the $P \to VV$ decays:

$$\begin{aligned}\frac{d^4\mathcal{P}(\vec{\omega},t)}{d\vec{\omega}\,dt} &\propto |A_0|^2 g_1(t) f_1(\vec{\omega}) + |A_{||}|^2 g_2(t) f_2(\vec{\omega}) \\ &+ |A_\perp|^2 g_3(t) f_3(\vec{\omega}) \pm Im(A_{||}^* A_\perp) g_4(t) f_4(\vec{\omega}) \\ &+ Re(A_0^* A_{||}) g_5(t) f_5(\vec{\omega}) \pm Im(A_0^* A_\perp) g_6(t) f_6(\vec{\omega}) \\ &\equiv \sum_{i=1}^{6} A_i g_i(t) f_i(\vec{\omega}) \end{aligned} \tag{1.77}$$

where A_i are self-defined (Latin index i runs from 1 to 6) and $f_i(\vec{\omega})$ are as follows:

$$f_1(\vec{\omega}) = \frac{9}{32\pi} 2\cos^2\psi(1-\sin^2\theta\cos^2\varphi) \tag{1.78}$$

$$f_2(\vec{\omega}) = \frac{9}{32\pi} \sin^2\psi(1-\sin^2\theta\sin^2\varphi) \tag{1.79}$$

$$f_3(\vec{\omega}) = \frac{9}{32\pi} \sin^2\psi\sin^2\theta \tag{1.80}$$

$$f_4(\vec{\omega}) = -\frac{9}{32\pi} \sin^2\psi\sin 2\theta\sin\varphi \tag{1.81}$$

$$f_5(\vec{\omega}) = \frac{9}{32\pi}\frac{1}{\sqrt{2}} \sin 2\psi\sin^2\theta\sin 2\varphi \tag{1.82}$$

$$f_6(\vec{\omega}) = \frac{9}{32\pi}\frac{1}{\sqrt{2}} \sin 2\psi\sin 2\theta\cos\varphi. \tag{1.83}$$

One should be careful not to confuse the amplitudes A_α and A_i; the latter are the bilinear forms of the amplitudes A_α. Equation 1.77 is valid for initially produced B ($\overline{B}$) mesons when the upper (lower) sign is taken for the interference terms (4 and 6) between the (C)P even and odd amplitudes. The time evolution of the angular terms, the $g_i(t)$'s, also depends on whether a B or a $\overline{B}$ is initially produced, and the mixing and decays that follow. Below we discuss this separately for the B_s and B_d decays under study.

As we have seen in Section 1.3.2, the CP eigenstates in the B_s system are very close to the mass eigenstates (also eigenstates of well-defined lifetime). The observed final states of the $B_s \to J/\psi\phi$ decay are CP eigenstates, thus to a very good approximation, the CP-even angular terms (involving $|A_{0,||}|^2$) develop in time with $e^{-\Gamma_L t}$, while the CP-odd angular term (involving $|A_\perp|^2$) develops with $e^{-\Gamma_H t}$. The results of a careful calculation [47] of the time evolution of each of the six angular terms in Equation 1.77 are given in Table 1.3. These include effects linear in the small[7] CP violating phase, $\delta\phi_s$, arising in the interference of decays with and without mixing.

[7] $\delta\phi_s = arg\left(\frac{V_{ts}V_{tb}^*}{V_{ts}^*V_{tb}}\frac{V_{cs}^*V_{cb}}{V_{cs}V_{cb}^*}\right)$ is zero to first order in Wolfenstein's parameterization. Taking into account higher-order terms gives $\delta\phi_s = 2\beta_s \approx 2\lambda^2\eta \approx 0.03$ [47].

i	A_i and time evolution, $g_i(t)$
1	$\|A_0\|^2\left[e^{-\Gamma_L t} \mp e^{-\Gamma t}\sin(\Delta M_s t)\delta\phi_s\right]$
2	$\|A_{\|\|}\|^2\left[e^{-\Gamma_L t} \mp e^{-\Gamma t}\sin(\Delta M_s t)\delta\phi_s\right]$
3	$\|A_\perp\|^2\left[e^{-\Gamma_H t} \pm e^{-\Gamma t}\sin(\Delta M_s t)\delta\phi_s\right]$
4	$\pm\|A_{\|\|}\|\|A_\perp\|\left[e^{-\Gamma t}\sin(\delta_\perp - \delta_{\|\|} - \Delta M_s t) \pm \frac{1}{2}[e^{-\Gamma_H t} - e^{-\Gamma_L t}]\cos(\delta_\perp - \delta_{\|\|})\delta\phi_s\right]$
5	$\|A_0\|\|A_{\|\|}\|\cos\delta_{\|\|}\left[e^{-\Gamma_L t} \mp e^{-\Gamma t}\sin(\Delta M_s t)\delta\phi_s\right]$
6	$\pm\|A_0\|\|A_\perp\|\left[e^{-\Gamma t}\sin(\delta_\perp - \Delta M_s t) \pm \frac{1}{2}[e^{-\Gamma_H t} - e^{-\Gamma_L t}]\cos(\delta_\perp)\delta\phi_s\right]$

Table 1.3: Time evolution of initially produced $B_s/\overline{B}_s$ mesons (upper/lower sign). Terms of the order of $(\delta\phi_s)^2$ are neglected.

It is easy to imagine that we cannot directly apply the results of Table 1.3 in our analysis. The two circumstances that would appear to hold us back are:

- We do not know what ΔM_s is.
- We do not know if the observed decay comes from an originally produced B_s or $\overline{B}_s$.

However, the B_s decay mode we observe, $J/\psi(\mu^+\mu^-)\phi(K^+K^-)$, is not self-tagging, that is we do not know whether it is a B_s or $\overline{B}_s$ at the time of decay. For each time t we can consider the sum of decays coming from the incoherent sources of B_s and $\overline{B}_s$ mesons. If we assume equal numbers of B_s and $\overline{B}_s$ are produced, one finds from Table 1.3 that statistical cancellation occurs for all the terms involving ΔM_s, except the ones proportional to $(e^{-\Gamma_H t} - e^{-\Gamma_L t})\delta\phi_s$, which are very small anyway. This statistical cancellation becomes more precise as the size of the signal gets larger.

With the above in mind, we adopt the following approach to the B_s analysis:

- Assume equal numbers of B_s and $\overline{B}_s$ in the B_s sample.
- Assume that $\delta\phi_s = 0$, that is, ignore the possibility of any sizable CP violation in in the interference of $B_s - \overline{B}_s$ mixing and decay.

With these assumptions a "reduced" time-dependent angular distribution is obtained by adding the forms for initially produced B_s and $\overline{B}_s$ mesons, causing cancellation of the terms 4 and 6 involving interference between the CP-even and CP-odd decays:

$$\begin{aligned}\frac{d^4\mathcal{P}(\vec{\omega},t)}{d\vec{\omega}\,dt} &\propto |A_0|^2 e^{-\Gamma_L t} f_1(\vec{\omega}) + |A_{\|\|}|^2 e^{-\Gamma_L t} f_2(\vec{\omega}) \\ &\quad + |A_\perp|^2 e^{-\Gamma_H t} f_3(\vec{\omega}) + Re(A_0^* A_{\|\|}) e^{-\Gamma_L t} f_5(\vec{\omega}) \qquad (1.84)\end{aligned}$$

With Equation 1.84 it is possible to extract all $|A_\alpha|$, Γ_L, Γ_H and $\delta_{\|\|}$ (up to $\pm$ ambiguity) from the B_s sample in hand. However, unlike for B_d which is discussed below, information on $\delta_\perp$ can not be extracted.

In the $B_d - \overline{B}_d$ system the mass eigenstates are not close to the CP eigenstates, as the results of Section 1.3.2 imply. Therefore, when omitting the oscillation terms, the P-even amplitudes evolve as $\cos^2\beta e^{-\Gamma_L t} + \sin^2\beta e^{-\Gamma_H t}$ and the P-odd amplitude as $\sin^2\beta e^{-\Gamma_L t} + \cos^2\beta e^{-\Gamma_H t}$ [48]. In principle, this reduces, but still leaves a sensitivity to the first order in $\Delta\Gamma_d$. However, the fact that we observe flavor-specific B_d decays, *i.e.* $K^{*0} \to K^+\pi^-$ and *not* $K^{*0} \to K^0_S\pi^0$, ensures [30] that all terms linear in $\Delta\Gamma_d$ cancel out. The time-dependence for each of the six angular terms reduces to that given by Equation 1.59, which we re-write here for convenience:

$$g_i(t) = \frac{1}{2}\Big[\cosh\Big(\frac{\Delta\Gamma_d t}{2}\Big) \pm \cos(\Delta \mathrm{M}_d t)\Big] e^{-\Gamma_d t}$$

with the upper(lower) sign for initially produced $B_d(\overline{B}_d)$ mesons.

Section 1.3.2 points out that $\Delta\Gamma_d/\Gamma_d$ is very small. Ignoring terms quadratic and higher order in $\Delta\Gamma_d$, the time dependence becomes $e^{-\Gamma_d t}\cos^2(\Delta \mathrm{M}_d t/2)$ for decays from an initially produced B_d and $e^{-\Gamma_d t}\sin^2(\Delta \mathrm{M}_d t/2)$ for an initially produced $\overline{B}_d$. Taking these results and summing over the initially produced B_d and $\overline{B}_d$, we obtain the following distribution for the observed $B_d \to J/\psi K^{*0}$ decays:

$$\begin{aligned}\frac{d^4\mathcal{P}(\vec{\omega}, t)}{d\vec{\omega}\, dt} \propto \Big[& |A_0|^2 f_1(\vec{\omega}) + |A_{\|}|^2 f_2(\vec{\omega}) \\ & + |A_{\perp}|^2 f_3(\vec{\omega}) \pm Im(A^*_{\|}A_{\perp}) f_4(\vec{\omega}) \\ & + Re(A^*_0 A_{\|}) f_5(\vec{\omega}) \pm Im(A^*_0 A_{\perp}) f_6(\vec{\omega})\Big] e^{-\Gamma_d t},\end{aligned} \tag{1.85}$$

where the upper(lower) is used for the flavor specific final state of the B_d ($\overline{B}_d$) with $K^+\pi^-$ ($K^-\pi^+$).

The angular distribution given by Equation 1.85 can be used to extract all $|A_\alpha|$ as well as both $\delta_{\|}$ and $\delta_{\perp}$ (and, of course, Γ_d). From SU(3) flavor symmetry it is expected that the decay amplitudes for the B_d and B_s should agree within reason.

1.5 Analysis overview

The ultimate goal of the analyses described in this dissertation is to measure the width difference between the mass eigenstates of the $B_s - \overline{B}_s$ system. However, on the way to the final result a number of other interesting measurements are accomplished. These include measurement of the average lifetimes for the $B_{u,d,s}$ mesons as well as the measurement of the angular amplitudes in the studied B_d and B_s decays. Several novel techniques and analysis steps that could be (in fact, many already are) carried over "as is" into other analyses are developed and documented.

In this section we briefly review the status of the pertinent B meson measurements and discuss our analysis strategy and methodology.

1.5.1 B lifetimes

The experimental state of affairs, illustrated by Table 1.4, is such that the only precisely measured lifetimes (and masses, for that matter) are those of the B_u and B_d, which are copiously produced at B factories. The other two ground state B mesons, B_s and B_c, are not nearly as abundant and therefore less well measured.

B_q	Charge, $\|e\|$	Mass, MeV/c^2	Lifetime, ps	Production fraction, %
B_d	0	5279.4 ± 0.5	1.537 ± 0.015	38.8 ± 1.3
B_u	+1	5279.0 ± 0.5	1.671 ± 0.018	38.8 ± 1.3
B_s	0	5366.0 ± 1.2	1.461 ± 0.057	10.6 ± 1.3
B_c	+1	6.4 ± 0.4	$0.46\ ^{+0.18}_{-0.16}$	< 1

Table 1.4: Properties of the ground state B mesons. Compiled from References [25,49]

In this dissertation we are only considering light-flavored B mesons: B_u, B_d, and B_s. B_c mesons are not included in our analysis for two reasons. Firstly, their production rate is so small – only about 1/40 of that of $B_{u,d}$ – that one needs much more data than presently available in order to achieve a precision measurement. Secondly, in addition to decaying via b quark disintegration, B_c can also decay via the CKM-favored $c \to Ws$ quark sub-process, and thus require a theoretical treatment substantially different from that of $B_{u,d,s}$.

We start off by measuring the (average) lifetimes of B_u and B_d mesons. At this time, given the statistics, we are not likely to challenge the precision achieved by B factories. However, measuring these lifetimes is not at all meaningless.

Firstly, an environment completely different from B factories requires a totally different approach. We re-work and improve the lifetime extraction method used by the CDF Collaboration during Run I [50], and demonstrate that the results obtained with this new method for $B_{u,d}$ mesons are in excellent agreement with B factory measurements.

The second important aspect of measuring any relatively short lifetimes at high energy hadron colliders is that the proper decay length resolution is much better (due to larger boosts and, consequently, longer flight distances) than those at B factories, translating into smaller statistical uncertainty even with smaller-size samples. Also, the systematic effects are to a large extent "orthogonal" to those at B factories. The results of the measurements presented here indicate that it is reasonable to expect that toward the end of Run II lifetime measurements from Tevatron will rival the precision of B factory measurements, which makes having a proven lifetime extraction method ever more important.

In what concerns B_s lifetime(s), the progress has so far been rather limited. The B_s samples previously available were not adequate to incisively confront the theory. This situation persisted largely due to the fact that the B_s is not produced in the decays of $\Upsilon(4S)$ resonance, and for that reason cannot be studied at the currently operating B factories. Samples obtained by LEP were also small because of cross-sections much smaller than Tevatron's, but comparable total integrated luminosity.

At hadron colliders the B cross-section is very large ($\sim 100\,\mu$b at upgraded Tevatron), however, B_s production fraction, f_s, is only about a quarter of B_d production fraction, f_d. Still the sample of exclusively reconstructed B_s mesons considered in this analysis is several times larger than any of its predecessors.

Apart from being so scarce, B_s mesons provide an additional challenge because of the expected lifetime difference between the two mass eigenstates in the $B_s - \overline{B}_s$ system, which effectively splits the sample into two different ones. To extract the lifetimes properly from any given sample not only needs one a way of measuring lifetimes, but also a way to know at least the relative fraction of heavy and light mass eigenstates decays in the sample.

1.5.2 Angular decay amplitudes in the $P \to VV$ B meson decays

The interest in angular amplitudes in the $P \to VV$ B meson decays started a few years back, when people realized the utility of $B_d \to J/\psi K^{*0}$ decay (with $K^{*0} \to K^0_S\pi^0$) in measuring the angle β of the Unitarity Triangle. Indeed, if the two CP eigenstates can be disentangled (or one CP eigenstate dominates) then this decay can be used to measure the angle β much the same way in which the decay to the CP-odd eigenstate $B_d \to J/\psi K^0_S$ is used. Also a number of phenomenological models, *e.g.* [51–53], based on the factorization hypothesis, make predictions as to what $|A_0|^2$ should be. The extent to which the factorization hypothesis itself is valid can be tested as well, because this hypothesis implies that phases of all three amplitudes are the same (modulo π). Deviations from this would imply significant final state interactions and breakdown of factorization.

The original measurements have been performed by CLEO and CDF collaborations, who later on updated their results [54, 55]. CDF has also included the results for $B_s \to J/\psi\phi$ as a further check of the factorization and a first experimental attempt to see how useful this decay can be for measuring $\Delta\Gamma_s/\Gamma_s$. The current world averages for $B_{u,d}$ are dominated by results coming from BaBar and Belle [56,57]. The summary of the above mentioned measurements is given in Table 1.5. It should be

$B \to J/\psi K^*$	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\delta_\parallel$	$\delta_\perp$
CLEO (1997)	$0.52 \pm 0.07 \pm 0.04$	$0.16 \pm 0.08 \pm 0.04$	$3.00 \pm 0.37 \pm 0.04$	$-0.11 \pm 0.46 \pm 0.03$
CDF (2000)	$0.59 \pm 0.06 \pm 0.01$	$0.13\ ^{+0.12}_{-0.09} \pm 0.06$	$2.16 \pm 0.46 \pm 0.13$	$-0.56 \pm 0.53 \pm 0.13$
BaBar (2001)	$0.60 \pm 0.03 \pm 0.02$	$0.16 \pm 0.03 \pm 0.03$	$2.50 \pm 0.20 \pm 0.08$	$-0.17 \pm 0.16 \pm 0.07$
Belle (2002)	$0.62 \pm 0.02 \pm 0.03$	$0.19 \pm 0.02 \pm 0.03$	$2.83 \pm 0.19 \pm 0.08$	$-0.09 \pm 0.13 \pm 0.06$

$B_s \to J/\psi\phi$	$\lvert C_0\rvert^2$	$\lvert C_\perp\rvert^2$	$\delta_\parallel$	$\delta_\perp$
CDF (2000)	$0.61 \pm 0.14 \pm 0.02$	$0.23 \pm 0.19 \pm 0.04$	$1.12 \pm 1.29 \pm 0.18$	—

Table 1.5: The results of the earlier measurements of the angular amplitudes in $B \to J/\psi\ K^*/\phi$ decays.

noted that B factories combine the amplitude results from several decay channels, including $B_u \to J/\psi K^{*+}$, under the assumption of isospin symmetry, which motivates

the generic notation $B \to J/\psi K^*$ in the above table. For $B_s \to J/\psi\phi$ in Table 1.5 we changed the notation for the "amplitudes" from A_α to C_α, because what was actually measured are the time-integrated fractions, rather than the amplitudes. The distinction between the two is explained in Appendix J.

With the present analysis we intend to achieve a precision in measuring angular decay amplitudes in $B \to J/\psi K^*$ similar to that of the latest published results by BaBar and Belle [56, 57] as well as to make a big step forward in measuring the amplitudes in $B_s \to J/\psi\phi$.

1.5.3 The status of and considerations for $\Delta\Gamma_s/\Gamma_s$

There have been a number of experimental attempts to make a statement about $\Delta\Gamma_s/\Gamma_s$ either with a limit or a number with uncertainty. However, all of these attempts suffered from very low statistics and in some cases relied on controversial assumptions. In the current section, while reviewing the earlier results we try to illuminate at least the most important issues that arise in $\Delta\Gamma_s/\Gamma_s$ measurements.

As a sensible benchmark number for $\Delta\Gamma_s/\Gamma_s$ we choose 0.12 ± 0.06 calculated in References [30, 45]. Different authors have arrived at different numbers, but most of them are in the general proximity of 0.12. The experimental results can be found on either side of 0.12 more often than not with large deviations, but still agreeing with the prediction because of large uncertainties (Table 1.6).

There are three principle methods of $\Delta\Gamma_s/\Gamma_s$ extraction that have been employed so far. We describe them briefly, starting from the most popular. Appendix N contains formulae potentially useful when reviewing the material of this section.

Method 1. This method relies on measuring a lifetime from semileptonic B_s decays (with a single-exponential fit) and then using HQE prediction [38]:

$$\left|\frac{\Gamma_s}{\Gamma_d} - 1\right| < 0.01. \tag{1.86}$$

Indeed, the semileptonic decays are flavor-specific, *i.e.* a B_s decay yields an l^+ and a $\overline{B}_s$ decay yields a l^-. At the same time, both B_s^L and B_s^H contain equal amount of flavor eigenstates B_s and $\overline{B}_s$ (Equations 1.47 and 1.64 – 1.66). The above two facts combined mean that semileptonic decay width is the same for B_s^L and B_s^H: $\Gamma_L^{semil.} = \Gamma_H^{semil.}$. However, the semileptonic branching fractions, $\Gamma_L^{semil.}/\Gamma_L$ and $\Gamma_H^{semil.}/\Gamma_H$, are different, because $\Gamma_L \neq \Gamma_H$. This difference in the branching fractions means that there is unequal proportion of heavy and light B_s mass eigenstate decays in the semileptonic sample. Namely, assuming that equal number of B_s^L and B_s^H are produced, the number of B_s^L decays is proportional to $1/\Gamma_L$ and the number of B_s^H decays is proportional to $1/\Gamma_H$. Basically, there is fewer B_s^L semileptonic decays because of the additional (those that give rise to $\Delta\Gamma_s = \Gamma_L - \Gamma_H > 0$) decay channels available for B_s^L. In terms of fractions there is

$$\frac{1/\Gamma_L}{1/\Gamma_L + 1/\Gamma_H} = \frac{\Gamma_H}{2\Gamma_s} \tag{1.87}$$

B_s^L decays and

$$\frac{1/\Gamma_H}{1/\Gamma_L + 1/\Gamma_H} = \frac{\Gamma_L}{2\Gamma_s} \tag{1.88}$$

B_s^H decays. Appendix A explains that a single lifetime fit on any sample returns simply the average lifetime of all events in the sample. Applied to a sample of semileptonic B_s decays it means that the single exponential fit should return

$$\tau_{B_s}^{semil.} = \frac{\Gamma_H}{2\Gamma_s}\tau_L + \frac{\Gamma_L}{2\Gamma_s}\tau_H = \frac{1}{\Gamma_s}\frac{1+\frac{1}{4}\frac{\Delta\Gamma_s^2}{\Gamma_s^2}}{1-\frac{1}{4}\frac{\Delta\Gamma_s^2}{\Gamma_s^2}}. \tag{1.89}$$

If one now takes $\Gamma_s = \Gamma_d$ ($\frac{1}{\Gamma_s} = \frac{1}{\Gamma_d} \equiv \tau_{B_d}$) from Equation 1.86 and a measured value of $\tau_{B_s}^{semil.}$, Equation 1.89 can be solved for $\Delta\Gamma_s/\Gamma_s$. We would like to emphasize, however, that the current central values of τ_{B_d} (Table 1.4) and $\tau_{B_s}^{semil.}$ (Table 1.6) are in *disagreement* with $\tau_{B_s}^{semil.} \geqslant \tau_{B_d}$ implied by Equation 1.89. This observation has been made by others [58].

It is hard to identify the source of the above problem. The prediction of Equation 1.86 has been independently confirmed and the experimental results are what they are, given the uncertainties. We would prefer not to delve into a discussion of whether it is $\tau_{B_s}^{semil.}$ that is too low, or Equation 1.86 that is not entirely accurate. For the purposes of this dissertation we choose not to rely on B_s semileptonic lifetime in any way, but we do investigate the utility of the constraint of Equation 1.86 in improving the $\Delta\Gamma_s/\Gamma_s$ sensitivity. We do clearly state which of the results of the current dissertation use this constraint and which ones do not.

Method 2. In this method the lifetime in the decay $B_s \to D_s^{(*)+}D_s^{(*)-}$ is measured. This decay is (almost) entirely CP-even, thus the extracted lifetime is the lifetime of the CP-even mass eigenstate, τ_L. In order to extract $\Delta\Gamma_s/\Gamma_s$ such measurement can then be combined[8] using the algebra from Appendix N with the semileptonic B_s lifetime [59], or with the prediction of Equation 1.86 [60].

Method 3. This method is different from the previous two in that that it is entirely self-contained. The drawback, which can be considered minor unless an extremely precise measurement of $\Delta\Gamma_s/\Gamma_s$ is desired, is that the method is based on a somewhat speculative assumption that the total $\Delta\Gamma_s$ is due to just one (set of) decay(s): $B_s \to D_s^{(*)+}D_s^{(*)-}$. Indeed, only the modes that are common to B_s and $\overline{B}_s$ (real intermediate states of the box diagrams) define Γ_{12} and, consequently, $\Delta\Gamma$. Out of these, the dominant ones are governed by $b \to c\bar{c}s$ transition, while $b \to c\bar{u}s$, $u\bar{c}s$, $u\bar{u}s$ are CKM-suppressed. Reference [42] provides a credible argument that $b \to c\bar{c}s$ is dominated by the CP-even $B_s \to D_s^{(*)+}D_s^{(*)-}$ decay, so that

$$\Delta\Gamma_s = \Gamma(B_s^{(L)} \to D_s^{(*)+}D_s^{(*)-}). \tag{1.90}$$

[8]The caveat from Method 1 still applies.

Method	Sample	Measurement	$\Delta\Gamma_s/\Gamma_s, \%$
1	$B_s \to D_s^+ l^- \bar{\nu}_l X$	$\tau_{B_s}^{semil.} = (1.46 \pm 0.07)$ ps	< 29
2	$B_s \to \phi\phi X$	$\tau_L = (1.27 \pm 0.33 \pm 0.07)$ ps	$45\ ^{+80}_{-49}$
3	$B_s \to \phi\phi X$	$Br(B_s^{(L)} \to D_s^{(*)+} D_s^{(*)-}) = (23 \pm 10\ ^{+19}_{-9})\%$	$26\ ^{+30}_{-15}$

Table 1.6: Status of the experimental knowledge of $\Delta\Gamma_s/\Gamma_s$. The upper limit is quoted at the 95% C.L. The results are collected from References [59, 60].

Now, taking into account that

$$\frac{\Delta\Gamma_s}{\Gamma_L} = \int_0^\infty \Gamma(B_s^{(L)} \to D_s^{(*)+} D_s^{(*)-}) \cdot e^{-\Gamma_L t} dt \equiv Br(B_s^{(L)} \to D_s^{(*)+} D_s^{(*)-}), \qquad (1.91)$$

we see that it suffices to measure the branching fraction $Br(B_s^{(L)} \to D_s^{(*)+} D_s^{(*)-})$ in order to determine $\Delta\Gamma_s/\Gamma_s$ from (using $\Gamma_L = \Gamma_s\left[1 + \frac{\Delta\Gamma_s}{2\Gamma_s}\right]$, Appendix N)

$$\frac{\frac{\Delta\Gamma_s}{\Gamma_s}}{1 + \frac{\Delta\Gamma_s}{2\Gamma_s}} = Br(B_s^{(L)} \to D_s^{(*)+} D_s^{(*)-}). \qquad (1.92)$$

In the present analysis we employ a method which is also self-sufficient, but may benefit from additional information if we choose to include it. As one can see from Section 1.4, the angular analysis in $B_s \to J/\psi\phi$ at least in principle can separate the CP-odd decays from CP-even ones, thereby separating decays of the two mass eigenstates, such that the two lifetimes, τ_L and τ_H, can be extracted. This method does not rely on any earlier measurements or involve any dubious theoretical models. The measurement is entirely driven by the data in the sample on which the measurement is performed.

At the same time, the method is flexible enough that additional information, *e.g.* Equation 1.86, can be used to improve the sensitivity. However, introduction of such additional information is, firstly, done in a clear, well controlled fashion and, secondly, not critical for getting the result.

1.5.4 Analysis strategy

In sections 1.5.1–1.5.3 we have reviewed the current status of $B_{u,d,s}$ lifetimes, $B_{d,s}$ angular decay amplitudes and $\Delta\Gamma_s/\Gamma_s$ — all of the quantities we intend to measure. Hopefully, that gives a sense of the sheer size as well as the complexity of the experimental project that we are undertaking. In order to be successful and stay on course one needs a well defined strategy and the right tools. The tools are numerous and far too complicated, so we defer the discussion of them to Chapter 3 and Appendices. In this section we summarize the "Analysis overview" as well as the entire introductory chapter by presenting the analysis strategy.

We begin by measuring the lifetime of B_u mesons reconstructed in the exclusive

decay $B_u \to J/\psi(\mu^+\mu^-)K^+$. This is the largest (over 3000 signal candidates) sample of B mesons available to us. The latter fact together with the fact that τ_{B_u} is fairly well known (Table 1.4) make this measurement not just a measurement, but also an important check of the lifetime extraction tools that we subsequently apply on the smaller samples of the other B meson species.

Assuming success in the first step, we proceed to measuring the (average) lifetimes of B_d and B_s mesons in their respective decays to $J/\psi(\mu^+\mu^-)K^{*0}(K^+\pi^-)$ (over 1000 signal candidates) and $J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ (few hundreds signal candidates). From the first measurement of angular decay amplitudes [55] and/or by SU(3) symmetry with B_d, we expect that CP-even component of $B_s \to J/\psi\phi$ is small, about 20%. Thus, we might see an indication of a sizable $\Delta\Gamma_s/\Gamma_s$ already at this step if the extracted average B_s lifetime turns out to be notably smaller than τ_{B_d}.

In the third step we perform the time-dependent angular analysis of the $B_d \to J/\psi K^{*0}$ decays, extracting angular decay amplitudes and, again, B_d lifetime, τ_{B_d}. The extracted amplitudes are cross-checked against previous measurements (Table 1.5).

Lastly, we modify the fitter used in step 3 such that the functional form given by Equation 1.84, not Equation 1.85, is being fit to the signal in the sample. The modified fitter is then applied to the B_s data in hand. The extracted angular decay amplitudes are compared to the earlier measurement (Table 1.5) and to the B_d results from step 3. In the latter comparison we expect semi-quantitative agreement on the grounds of SU(3) symmetry. If everything works, we have accomplished the first direct measurement of $\Delta\Gamma_s/\Gamma_s$, independent of previous B_s lifetime measurements!

Chapter 2

Experimental apparatus and getting the data

The B mesons of interest are produced as a result of head-on proton–antiproton collisions. A detector is used to collect information about the interactions taking place in these collisions. This chapter describes the two major components crucial for the study we are undertaking: the Fermilab Tevatron, which provides $p\bar{p}$ collisions, and the CDF II detector allowing observation of the final state particles from which the global picture of the interactions is reconstructed. Some low level procedures, *e.g.* tracking and vertexing, as well as data manipulations, such as data set creation, which precede a typical analysis are also discussed in this chapter.

2.1 Tevatron

Fermilab is home to a very sophisticated accelerator complex a simplified drawing of which is shown in Figure 2-1. One of the things this complex is capable of providing is $p\bar{p}$ collisions at $\sqrt{s} = 1.96\,\text{TeV}$. This involves several stages, such as: preparing protons and antiprotons, injecting them into the Tevatron, accelerating the two beams up to the energy of 980 GeV and colliding them at the interaction regions. These steps of the process are described below along with the accelerator systems involved.

2.1.1 Proton source: Cockroft-Walton chamber, Linac, and Booster

The process leading to the $p\bar{p}$ collisions begins with electrical discharges into a hydrogen gas contained in a Cockroft-Walton chamber, across which there is a large electrostatic field. The discharges produce negative ions, which are accelerated by a positive voltage applied to a so-called extractor plate. The ions are then fed through a magnetic transport system, which filters out non-H^- species and delivers the H^- ions to a linear accelerator, the Linac.

The Linac [63] picks up the H^- ions at the energy of 750 KeV, accelerates them, and injects them in the Booster. The first half-length of the Linac is the Drift Tube

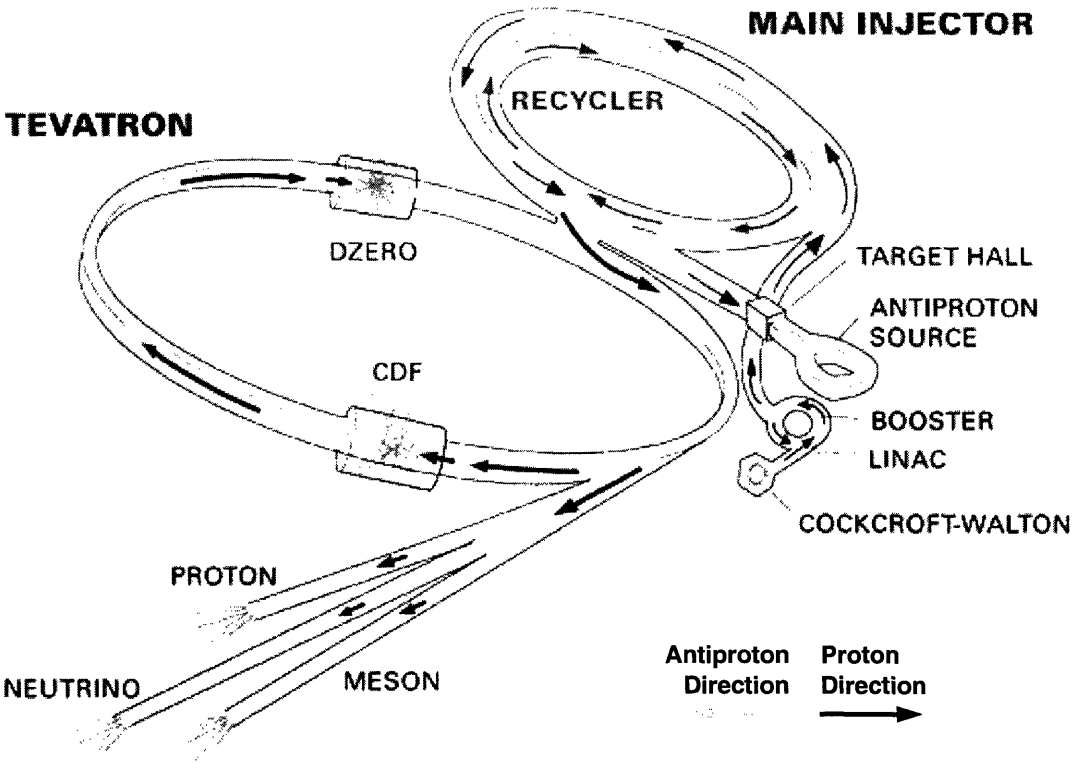


Figure 2-1: Fermilab's accelerator chain.

Linac, which accelerates the ions to 116 MeV. The second half of the (old-fashioned) Drift Tube Linac has been replaced with a Side-coupled Cavity Linac, allowing ions to achieve an ultimate energy of 400 MeV.

The Booster [62, 64] is a rapid cycling synchrotron about 150 meters in diameter. It uses a loading scheme that overlays injected beam with the one already circulating in the machine. One may argue that this violates the Liouville's Theorem[1]. This, in fact, would have been the case with the original H^+ source, which produced even larger beam current than the present H^- source, but was limited to injection over a single Booster turn. The H^- source and the overlaid beam injection were developed to increase beam intensity from the Booster by working around Liouville's theorem. The idea exploits the fact that it is a proton (H^+) beam that circulates in the Booster, while the incoming Linac beam is a beam of H^- ions. Superimposing H^- and H^+ beams constitutes no violation of the Liouville's Theorem. Once it is accomplished, the mixed beams go through a carbon foil, which strips off the electrons turning the H^- ions into protons; to minimize the beam blow-up the beams are made pass through the foil only during loading. As a result one gets an increased intensity proton beam confined to virtually the same phase space volume of the original machine with single turn loading.

[1]The density in phase space remains constant along a dynamic trajectory (in the presence of only conservative forces), in particular two dynamic trajectories cannot merge.

The revolution period in the Booster at injection is 2.22 ms, while the pulse length in the Linac is approximately 40 ms. A portion of the Linac beam is selected by a 400 MeV chopper magnet, while the remainder is sent to one of the Linac dumps. By extending the chop length, multiple Booster turns are generated. Although the Linac beam pulse is long enough for more than 18 Booster "turns", operationally 10–12 turns is usually the practical limit for maximum intensity.

Once the bare protons are collected in the Booster, they are accelerated to an energy of 8 GeV by the conventional method of varying the phase of the RF fields in the accelerating cavities [65] and subsequently injected into the Main Injector. One full Booster "batch" contains a maximum of 5×10^{12} protons, which sets a convenient unit for the number of protons throughout the accelerator complex. These protons are divided among 84 bunches spaced by 18.9 ns, which sets another convenient sub-unit, 6×10^{10} protons.

2.1.2 Main Injector

Before Run II, the protons and antiprotons in the Tevatron had been injected by the Main Ring, which was the original 400 GeV proton synchrotron built during the early 1970's. The Main Ring was not designed as an injector to the Tevatron, it was only later adapted to suit this purpose, which imposed a number of limitations on its performance. With significant improvement in antiproton production capability as the prized goal, Fermilab upgraded its accelerator complex with the Main Injector (MI) [62]. The Main Injector is a new rapid cycling accelerator designed to do the following:

- accept 8 GeV protons or antiprotons from
  - the Booster,
  - the antiproton Accumulator,
  - the Recycler,
- accelerate protons to 120 GeV and deliver them to
  - the antiproton production target,
  - fixed target area,
  - NuMI beamline (for neutrino production),
- accelerate protons or antiprotons to 150 GeV and inject them into the Tevatron.
- accept 150 GeV antiprotons from the Tevatron and decelerate them to 8 GeV for transfer to the Recycler.

The MI is exactly seven times the circumference of the Booster. Allowing one empty "slot" for switching, it can hold six Booster batches with 84 bunches in each. In Collider Mode six bunches containing 6×10^{10} protons each are accelerated to *flat top* (nominal maximum energy for a given machine at which particles can be kept

for an extended time) of 150 GeV. At flat top the bunches are coalesced into a single bunch of 27×10^{10} protons (see Table 2.1 in Section 2.1.4) which is then injected into the Tevatron. The above constitutes a single MI cycle. Thirty-six such cycles are needed to fill the Tevatron with protons. As one may have noticed coalescing is not a 100% efficient process. Provided good coalescing efficiency is achieved, two, three or four proton bunches can be prepared during a single MI cycle, requiring correspondingly 18, 12 or 9 cycles to fill the Tevatron.

Once the Tevatron is filled with protons the MI goes back to its other duties which in fact take most of its time: antiproton production, which we discuss in Section 2.1.3, and super-intense beam delivery for fixed target / neutrino experiments. The operational possibilities are sufficiently flexible to interleave antiproton production and fixed target physics with only modest impact on either.

2.1.3 Antiproton source: Debuncher, Accumulator, and Recycler

The largest bottle-neck in a $p\bar{p}$ collider is the supply of the antiprotons. Being antimatter they are much harder to come by on Earth, so one has to be very inventive in order to obtain a significant number of them.

For efficient acceleration and system-to-system transfer of particles, and ultimately for the highest luminosity, one wants the majority of the particles (antiprotons in this case) to be confined to a small phase space volume. This condition thermodynamically corresponds to a low temperature, which gives rise to a concept of a *cold* beam and *cooling*, *i.e.* producing a beam of particles confined to a small phase space volume.

Different cooling techniques exist. The *stochastic cooling* is a feedback based method. As particles orbit the storage ring they go past pickups and kickers, which are pairs of electrodes placed around the beam. Differential pickup signals (difference between the signals from the paired electrodes is proportional to particle's deviation from the nominal orbit) are used to produce the error signal. The error signal is then used by the kickers, which apply small momentum kicks effectively damping the amplitude of the oscillation around the nominal trajectory. Just described is the *transverse* or *betatron stochastic cooling*. In the *longitudinal* or *momentum stochastic cooling* energy spread of the beam is reduced by driving particles in the beam distribution toward a central momentum. In a momentum cooling system, the signals from the pickup electrodes are combined in the sum mode and similarly, the error signal is applied to the kicker electrodes, providing longitudinal fields to accelerate or decelerate the passing particles.

The *electron cooling* works on the principle of a heat exchanger. In this method a beam of electrons with a very small transverse kinetic energy is passed collinearly with the beam of particles being cooled (those that have relatively large transverse kinetic energy). Both beams travel at approximately the same velocity and as they interact, the transverse kinetic energy of the warmer beam is transferred to the electron beam, which is then collected at the end of the cooling session.

In either cooling method there is no violation of Liouville's theorem, because phase

space density is increased via application of non-conservative (friction) forces.

A very good introduction to cooling techniques, particularly those used at Fermilab, is given in Reference [66].

As we have seen in the previous section fairly high proton density is achieved in the Booster without any cooling. The sheer number of protons allows one to produce a cold beam just by throwing away protons that are outside of the desired phase space volume. However, for antiprotons, which are not so readily available, we do not have this luxury. Therefore, a distinctive effort in optimizing the production mechanism, and application of cooling is needed in order to obtain cold antiproton beam without discarding any significant number of antiprotons.

As mentioned in Section 2.1.2, antiproton production is one of the Main Injector's major tasks. In this mode of operation the MI accelerates a single Booster batch of 5×10^{12} protons (84 bunches) to 120 GeV and sends them to hit a nickel target. Just before extraction to the target RF manipulations are performed on the proton beam. They narrow the bunches in time at the expense of increasing momentum spread ($\Delta p/p$). This process is also known as bunch rotation[2]. It turns out that the $\Delta p/p$ of the antiprotons produced is minimally affected by the $\Delta p/p$ of the incident protons. At the same time, by narrowing the proton bunches prior to striking the target, the phase space density of the antiprotons is maximized.

When the beam strikes the production target a shower of secondary particles is produced. The large portion of the resulting cone of particles is rendered parallel by means of a lithium lens [66] (a cylindrical piece of lithium through which an axial current is run). The bunch structure of the beam coming off of the target is the same as that of the primary beam. A pulsed dipole magnet is used to select all negatively charged particles of approximately 8 GeV kinetic energy, which are subsequently injected into the Debuncher [62,66]. There the momentum spread of the 8 GeV secondaries is reduced through bunch rotation – the inverse of the procedure mentioned one paragraph above – and adiabatic de-bunching [66]. Both betatron and momentum stochastic cooling techniques [66, 68] are applied to further reduce the beam size and momentum spread.

Just before the next pulse arrives from the target (the cycle takes 1.5 seconds) the cooled antiprotons are transferred to the Accumulator [62, 66]. The purpose of the Accumulator, as its name implies, is to accumulate the antiprotons. Also a much more thorough cooling is performed. There are two ranges of orbits in the Accumulator. The inner orbits are referred to as the stack core, and the outer ones as stack tail. The injected antiprotons are moved into the stack tail, where momentum cooling acts decelerating them toward the stack core, which has a smaller orbit. After approximately 30 minutes, the antiprotons in the stack tail will have been decelerated into the domain of the core cooling systems. Three stochastic cooling systems act on the core during stacking, minimizing the *emittances*[3]. This process continues for

[2]One might think of a bunch as an ellipse in (t,p) space elongated along the t axis. After bunch rotation is applied it becomes an ellipse stretched along the p axis and narrowed along the t axis.

[3]Emittance is defined as the effective area of the beam in phase space and given by $x'\cdot dx$, where x' is the divergence. There are three emittances which are related to the phase space volumes in each

hours with the stack growing in size until the maximum Accumulator intensity is reached or the Tevatron needs to be refilled.

When the transfer of antiprotons to the MI is desired a portion of the beam core in the Accumulator is captured in four *buckets*[4] and is moved to the extraction orbit, where bunches suitable for capture by MI RF are prepared. Once extracted from the Accumulator and injected into the MI, antiprotons undergo acceleration from 8 GeV to 150 GeV and subsequent injection into the Tevatron. Nine cycles are needed to fill Tevatron with 36 antiproton bunches with 33×10^9 particles in each (see Table 2.1 in Section 2.1.4).

One other thing which should be mentioned is the Recycler ring [62,67]. Without it, the precious antiprotons left at the end of a collider *store* ($\sim$ 15 hour period of time when the colliding beams are retained in the machine) must be thrown away. The Recycler will make it possible to recover these antiprotons and re-use them in a later store. The Recycler will also take up the role of the Accumulator as the final storage for 8 GeV antiprotons allowing the existing Antiproton Source to perform more efficiently and produce more antiprotons per hour. Indeed, the Accumulator core size is probably limited to 2×10^{11} antiprotons, because the efficacy of stochastic cooling systems drops as $\frac{1}{N}$, where N is proportional to the linear charge density. The Recycler has a circumference 7 times that of the Accumulator, therefore for a given core size, line charge density is smaller and cooling capability is restored. In addition, electron cooling [69] is foreseen for the Recycler, in which case cooling rate does not scale with $\frac{1}{N}$.

The Recycler was included in the Fermilab Program in the spring of 1997 as an addition to the Main Injector project. Most of the lattice elements (dipoles and quadrupoles) are made out of permanent magnets and the ring shares the tunnel with the MI, thus operational and constructional costs are comparatively low, but the benefit is expected to be substantial.

2.1.4 Tevatron – the global picture

The Tevatron [62] is the highest energy hadron collider in the world. It accelerates beams of protons and antiprotons to the energy of 980 GeV, providing a center of mass (CM) energy of 1.96 TeV. It is also the first accelerator to employ superconducting magnets throughout, which required cryogenic cooling and, consequently, large scale production and distribution of liquid helium. Use of the superconducting elements is a necessity when achieving such high values of CM energy, but the added complexity has a benefit of reducing electricity consumption and hence the cost of operation.

---

dimension occupied by the particles in the beam. These are often reduced to transverse (measured in π-mm-mrad) and longitudinal(measured in eV-sec) emittances.

[4]A bucket (not to be confused with a *bunch*) is an accelerator physics term meaning a stable phase space region where beam may be captured and accelerated by RF cavities. The bucket width gives the maximum phase (or timing) range at the RF cavity which a particle may have and still complete the whole acceleration cycle. The bucket height is the corresponding limit on momentum spread. A group of particles captured in a bucket forms a bunch. A bucket may or may not contain a bunch.

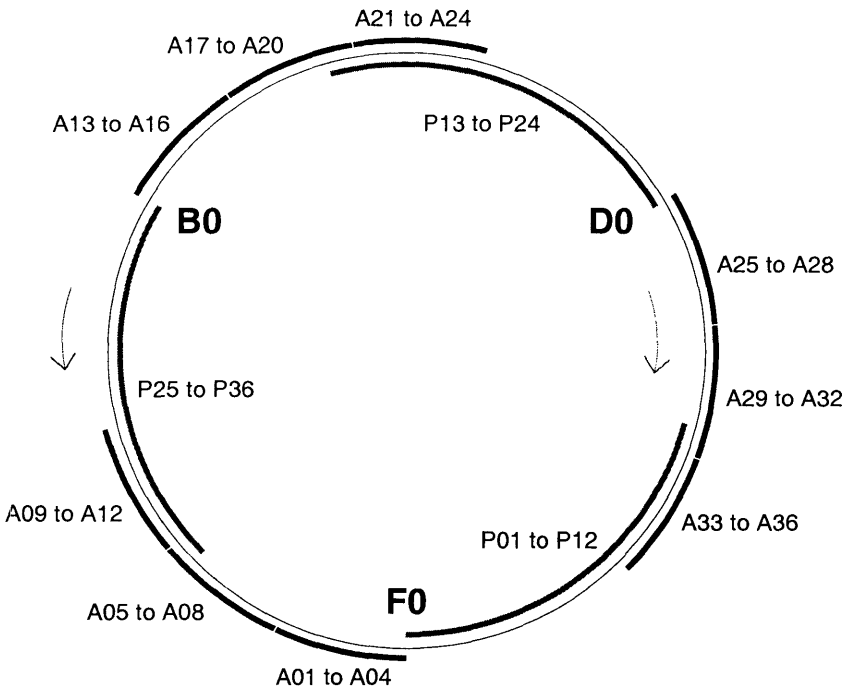


Figure 2-2: Beam spacing and injection configuration. Shown are three proton (clockwise) and antiproton (counterclockwise) bunch trains. Each contains 12 bunches spaced by 21 RF buckets (spacing is not shown). The proton and antiproton bunches are labeled P01, P02, ... and A01, A02, ... starting from the upstream end of the train so that A01 and P01 collide at F0.

The Tevatron is exactly 2 km in diameter. During Run IIa the Tevatron is operated in a 36 × 36 mode, which refers to the number of bunches in each of the two beams. The antiprotons are injected after the protons have already been loaded. Just before the antiproton injection a set of electrostatic separators are used to create a pair of non-intersecting helical closed orbits with the protons circulating on one strand of the helix and the antiprotons circulating on the other.

The beam configuration and injection scheme are illustrated by Figure 2-2. The standard 36 × 36 filling scheme consists of pattern of 12 bunches spaced by 21 RF buckets (396 ns). Each bunch "train" is followed by a 139-bucket gap called *abort gap*. The spacing of the antiproton bunch ensemble is the mirror image of the proton spacing.

As mentioned above, proton injection happens first. It follows a simple order: P01, P02, *etc.* The injection is accomplished via the low level RF transfer system and a kicker magnet with a fast rise time (396 ns). The antiprotons must be injected during the time that the proton abort gap passes through the kicker, so that the protons are not disturbed by the kicker's magnetic field. The only way this condition can be achieved is by rotating (or *cogging*) the antiproton distribution relative to the proton distribution for the injection of the various batches of antiprotons. Because there are three abort gaps in the proton beam, it is possible to inject three groups of antiprotons for each value of cogging. First, one third of each of the three antiproton

	Run IIa goal	08/2002	Main problems
$\bar{p}$ production rate, $\times 10^{10}$/hr	18	11	1) Larger than expected
Accumulator core size, $\times 10^{10}$	165	120	transverse emittance in the
$\bar{p}$ transfer efficiency, %	80	37	Accumulator
# p per bunch, $\times 10^{10}$	27	18	2) Long range beam-beam
# $\bar{p}$ per bunch, $\times 10^{9}$	33	12	interaction in the Tevatron
			reduces $\bar{p}$ lifetime, causes
Peak lumi., $\times 10^{31}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$	8.6	2.0	excessive emittance growth

Table 2.1: Summary of the performance of Fermilab accelerator complex.

trains is injected (A01-A04, A13-A16 and A25-A28). After that the antiprotons are cogged by 84 buckets, and again after the injection of A05-A08, A17-A20 and A29-A32. Prior to acceleration the antiprotons are rotated to the nominal collision point cogging.

Once the Tevatron loading is complete beams are accelerated to the maximum energy and collisions are initiated. With 36 × 36, there are 72 regions along the ring where (long range) bunch crossings occur. While 70 of these are parasitic, in the vicinity of B0 and D0 regions, surrounded by CDF and DØ detectors, a special effort is made to maximize the chance of a proton striking an antiproton. This is achieved with additional focusing and beam steering. Focusing, performed with quadrupole magnets, reduces the beam spot size and thus increases the *luminosity*. Indeed, instantaneous luminosity, to which the number of collisions per unit of time is proportional, to an approximation is given by:

$$\mathcal{L} = \frac{N_B N_{\bar{p}} N_p f}{2\pi(\sigma_{\bar{p}}^2 + \sigma_p^2)}, \tag{2.1}$$

where N_B is the number of bunches, $N_{(\bar{p}),p}$ is the number of the (anti)protons in a bunch, f is the bunch revolution frequency, and $\sigma_{(\bar{p}),p}$ is the effective width of the (anti)proton beam. Making $\sigma_{(\bar{p}),p}$ smaller and $N_{(\bar{p}),p}$ larger results in larger rate of collisions.

The instantaneous luminosity degrades over time as particles are lost and beams begin to heat up, mostly due to the long range beam-beam interaction. In the meanwhile more and more antiprotons are being stored in the Accumulator. After a period of time of about 15 hours, which is approximately one "luminosity lifetime", it becomes beneficial to terminate the current store and start a new cycle.

Table 2.1 summarizes Fermilab accelerator complex performance by listing Run IIa goals and current (as of this writing, 08/2002) performance characteristics.

2.2 The CDF II detector

CDF II is a general purpose solenoidal detector positioned around the B0 interaction region as shown in Figure 2-1. The layout is standard for detectors of this type. From

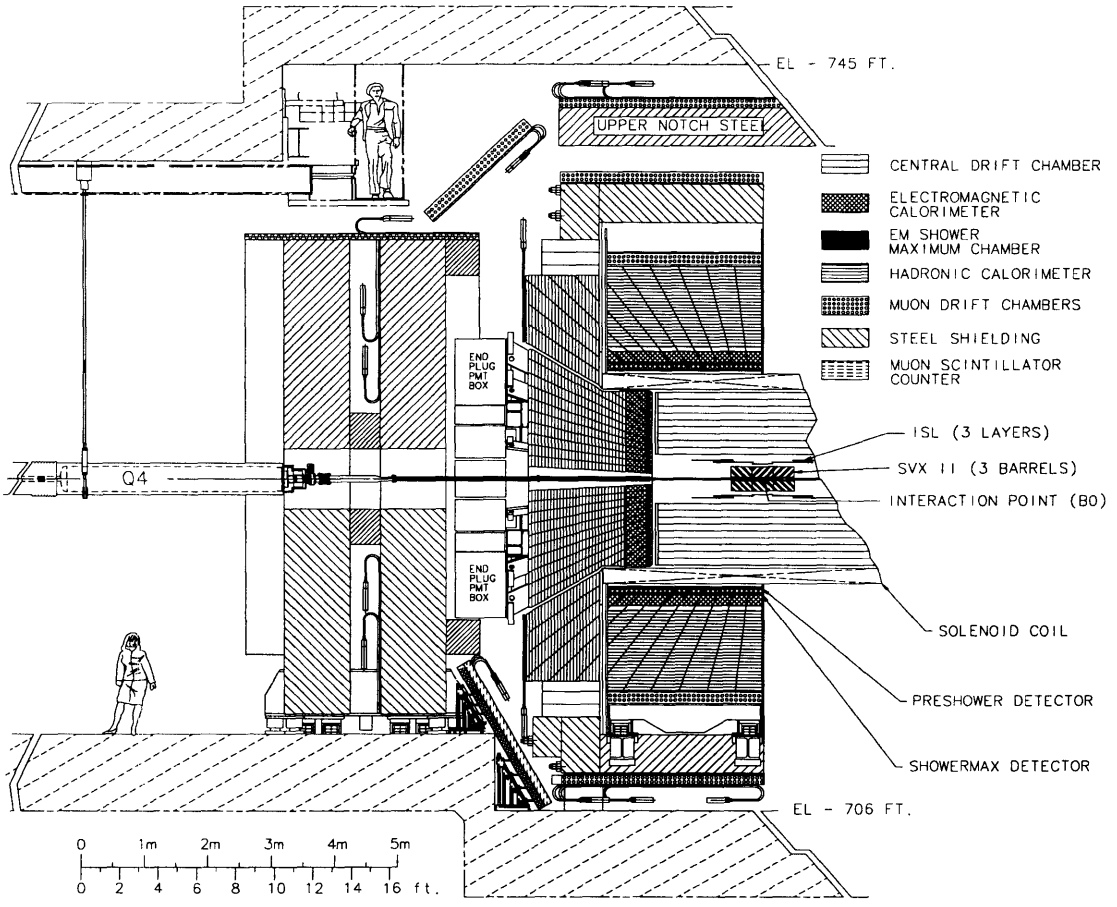


Figure 2-3: The CDF II detector.

the inside out there are: tracking system, magnet, electromagnetic (EM) calorimetry, hadron calorimetry, and muon detectors on the far outside. There are also systems, such as Time-of-Flight system and Cherenkov Luminosity Counter, which are not as universal as those just mentioned, but still are extremely important for particular classes of analyses. The elevation view of CDF II is shown in Figure 2-3.

The Cartesian coordinate system associated with CDF II has the detector's geometric center as its origin. The x axis is in the (horizontal) plane of the accelerator ring, pointing radially outward, and the y axis points vertically up. The z axis is chosen so as to complete the right-handed coordinate system. Beams travel through the detector approximately parallel to its z axis with protons moving in the the positive z direction. The center of the detector roughly coincides with the center of the beam crossing region.

The detector is roughly cylindrically and forward-backward symmetric, which makes it often convenient to work in cylindrical (z, r, ϕ) or polar (r, θ, ϕ) coordinates. The z axis is the same as for Cartesian system. The azimuthal angle ϕ runs in the transverse ($x-y$) plane, with $\phi = 0$ being the positive direction of the x axis. Polar angle θ is counted from the positive direction of the z axis.

Yet another set of coordinates replaces the angle θ in polar coordinates with the quantity

$$\eta = -\log\tan\frac{\theta}{2}, \tag{2.2}$$

called pseudo-rapidity. The choice of (r, η, ϕ) is convenient because it is defined by the symmetries of the experiment. Indeed, physics processes are invariant under rotation around the (unpolarized) beam, thus making ϕ a natural choice. To understand why η is a useful quantity one has to recall that the important feature of the proton–antiproton collisions is that it is *partons* (valence or sea quarks, or gluons) that are involved in fundamental interactions. They carry only a fraction of (anti)proton momentum, determined by the parton distribution function [73]. As a consequence of the possible imbalance in the longitudinal components of the momenta of the two interacting partons, the observed physics interactions often have large boosts along the z axis[5]. The quantity

$$\zeta \equiv \frac{1}{2}\log\frac{E+p_z}{E-p_z} \tag{2.3}$$

is called the rapidity. Under a boost β along the z axis it transforms as $\zeta' = \zeta + \tanh^{-1}\beta$, which means ζ intervals are invariant under such boosts. η is the ultra-relativistic/massless limit of ζ, in which the total energy E in the formula above can be replaced with the total momentum p:

$$\eta \equiv \frac{1}{2}\log\frac{p+p_z}{p-p_z} = -\log\tan\frac{\theta}{2}. \tag{2.4}$$

We see that ζ has reduced to a purely geometrical quantity η, appropriate to use as detector coordinate. At the same time η intervals are invariant under Lorentz boosts along z axis (provided that E can be approximated with p).

Detector components (wherever appropriate) are uniformly segmented in η and ϕ, which often allows one to simplify the analysis and avoid artifacts in certain kinematic distributions.

In the following sub-sections we give a general description of the CDF II detector with an emphasis on the elements relevant to the analyses described in this dissertation. A detailed description can be found in Reference [61]. Throughout this document we omit the "II" part of "CDF II" and "SVX II". There should be no ambiguity, because we always speak of the Run II incarnation of either, unless the opposite is explicitly stated.

2.2.1 Tracking and vertexing

Charged particles cause ionization as they pass through matter. Typically, this ionization is localized near the trajectory of the particle in little clusters called *hits*. Once detected, hits can be used to reconstruct particle's trajectory in the process known

[5]Same thing happens in the transverse direction, but because the transverse boosts of a(n anti)proton are tiny the fundamental interactions have negligible transverse motion.

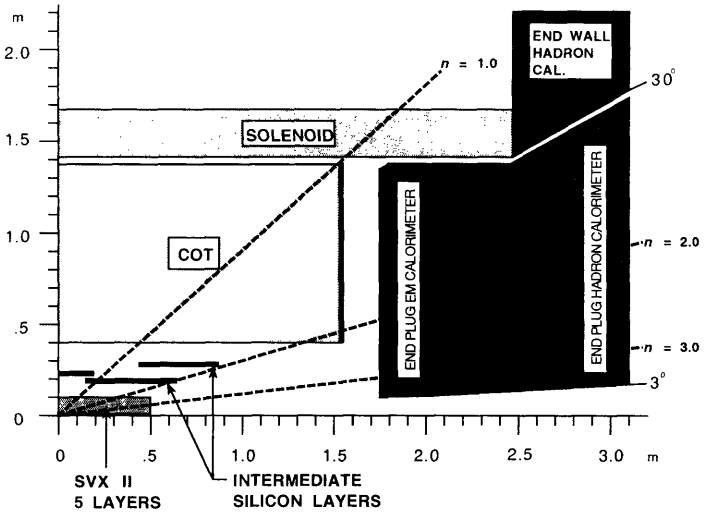


Figure 2-4: The $r-z$ view of the CDF tracking system.

as *tracking.*

In the upgrade for the Run II a great deal of effort was spent on increasing tracking coverage. The $r-z$ map of tracking systems is shown in Figure 2-4. We refer to this picture in the following sections when individual tracking systems are discussed.

The CDF tracking volume is permeated by a uniform magnetic field of magnitude $B = 1.4\text{T}$, directed along the z axis. In this environment charged particle trajectories are helices, which are described by five parameters:

z_0 – z coordinate of the point of the closest approach to the z axis,

d_0 – impact parameter (distance from the point of the closest approach to the z axis),

ϕ_0 – ϕ direction of the transverse momentum of the particle (tangential to the helix) at the point of the closest approach to the z axis,

$\cot\theta$ – helix pitch, defined as the ratio of the helix step to its diameter,

C – helix curvature.

Curvature and impact parameter are actually signed quantities as defined by

$$C = \frac{q}{2R}, \tag{2.5}$$

$$d_0 = q(\sqrt{x_c^2 + y_c^2} - R), \tag{2.6}$$

where q is the charge of the particle, (x_c, y_c) is the center of the helix as projected onto the $x-y$ plane, and R is its radius. These, as well as ϕ_0, are illustrated in Figure 2-5.

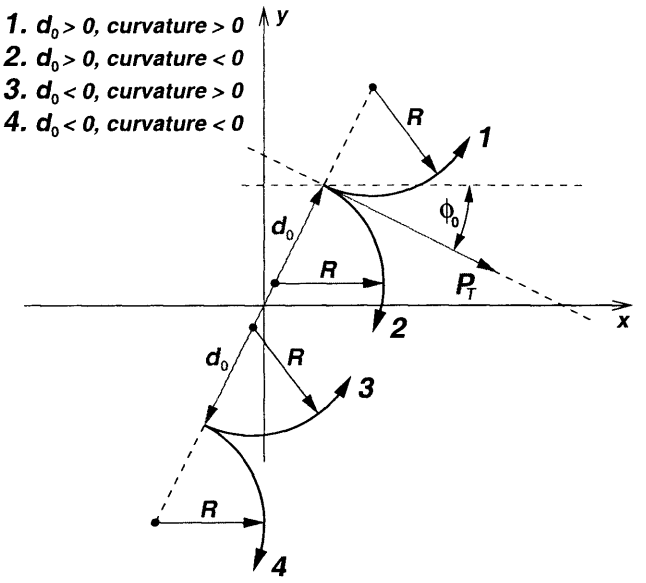


Figure 2-5: Illustration for track helix parameterization. Beamline is assumed to coincide with the z axis (perpendicular to the plane of the picture passing through (0,0)). In the transverse plane helices look like circles; radius R of the circle defines the track curvature C. Impact parameter d_0 is measured from the beamline. Shown are four identical tracks save for the signs of their curvatures and impact parameters.

From the helix parameters one can easily derive particle's transverse and longitudinal momenta:

$$p_T = \frac{c\,B}{2|C|} \tag{2.7}$$

$$p_z = p_T \cot\theta \tag{2.8}$$

In order to determine the track parameters a track fit is performed. This is basically a helical fit to a set of spatially distributed hits preselected according to the pattern recognition algorithm. Each hit going into the fit is ascribed an uncertainty universally determined in dedicated studies. The output of the fit is a set of track parameters and the error matrix, including inter-parameter correlations. We continue to discuss track fit later in this section after the two major tracking devices are introduced. It would be more appropriate to introduce the concept of vertexing first, however.

A particle's point of origin cannot be inferred from a helix alone, all we know is that it is somewhere on the helix. In many cases a particle's origin (production vertex) is found by intersecting the helix with at least one other helix corresponding to a particle which we believe has come from the same space point, *e.g.* as a result of a decay of a common parent particle. This process is referred to as *vertexing*. At CDF II vertexing is done using `C++` wrapper called `VertexFit` [74] of the Run I kinematic fitting `Fortran` code `CTVMFT` [75]. `VertexFit` determines the spatial coordinates of the decay point as well as their uncertainties. Determining the mass of the parent

particle and the corresponding uncertainty (from track parameters of the daughters and their masses) is also among its extensive capabilities.

There are two main systems that are the foundation for precise measurement of the helix parameters. These are: the Central Outer Tracker (COT) and the Silicon VerteX Detector (SVX), both of which we describe in detail below. There are also two additional systems: the Intermediate Silicon Layers (ISL) and Layer 00 (L00), which we describe only briefly.

Central Outer Tracker

In the region of $|z| < 155\,\text{cm}$ and between the radii of 44 and 132 cm tracking is done by the Central Outer Tracker, COT. The COT [61] is a cylindrical multi-wire open-cell drift chamber with 50:50 Ar-Ethane gas mixture as the active medium. Charged particles that travel through the gas mixture leave a trail of ionization electrons. These electrons drift toward the sense wires by virtue of the electric field created by cathode field panels and potential wires of the cell that they are in (Figure 2-6). The drift is not along the electric field direction because of the magnetic field along the z axis. In such crossed fields electrons originally at rest move in the plane perpendicular to the magnetic field at an angle α with respect to the electric field lines. The value of α, also known as a Lorentz angle, depends on the magnitude of both fields and the properties of the gas mixture. In the COT $\alpha \simeq 35°$.

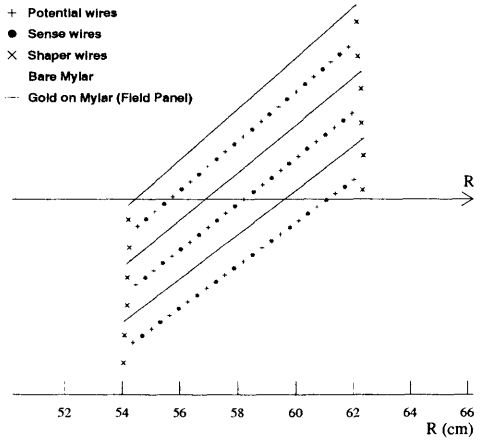


Figure 2-6: Transverse view of three cells from the COT superlayer 2. The arrow shows the radial direction or, equivalently, a very high p_T track trajectory. The electric field is roughly perpendicular to the field panels; the drift velocity would be vertical in this picture (see text). The angle between wire-plane of the central cell and the radial direction is 35°.

Resolution-wise the optimal situation is when the drift direction is perpendicular to that of the track. Usually the optimization is done for high p_T tracks, which are almost radial. Therefore all cells are tilted by 35° away from the radial so that ionization electrons drift in the ϕ direction. When the electrons get near the sense wires the local $\frac{1}{r}$ field vigorously accelerates them causing more ionization. These

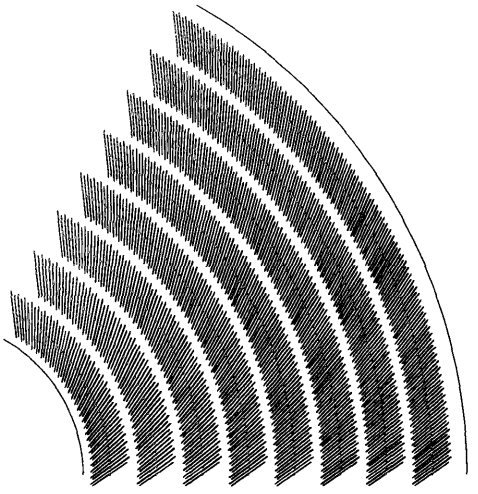


Figure 2-7: 1/6th of the COT endplate. Shown are the wire-plane slots grouped into eight superlayers. Each slot is tilted by 35° degree with respect to the radial direction. In the stereo superlayers cells are strung using an eight slot offset between slots on opposing endplates.

secondary ionization electrons form an "avalanche" producing a signal (hit) on the sense wire that is further amplified and shaped by the electronics attached to the end of the wire. The $r-\phi$ position of the track with respect to the sense wire is then inferred from the arrival time of the signal.

There are at least two other advantages to the wire-plane tilt. First, the tilt allows wire-planes to overlap in the radial view, which means that high p_T tracks must pass very close to at least one wire in each superlayer. This is taken advantage of by the eXtremely Fast Tracker (XFT) in the Level-1 trigger, as described in Section 2.3.1. Also the large tilt provides a lever in resolving the left/right ambiguity. Indeed, a particle passing by a wire-plane leaves several hits, which are grouped into a track segment, but since we have no way of knowing whether an individual hit is on the left of the respective wire or on the right of it there are actually two segments, one being the mirror image of the other. The angle between the two track segments – $\tan^{-1}(2 \cdot \tan\alpha) \approx 54°$ – is large, which renders pattern recognition problem much easier.

A charged particle that traverses the entire COT in the radial direction ideally has 96 position measurements. The cells – each includes 12 sense wires – are arranged in eight radially spaced superlayers (see Figure 2-7 and Table 2.2). The superlayers alternate between axial alignment (with cell plane parallel to the z axis) and small angle stereo alignment (with cell planes rotated by approximately $\pm 2°$ about radial direction, $r|d\phi/dz| \approx 0.05$).

Track reconstruction in the COT begins from finding matching track segments or just individual hits in the axial superlayers [76–78]. Once the $r-\phi$ projection of the track is reconstructed we get the measurement of C, d_0 and ϕ_0 helix parameters. When looked at in the $r-\phi$ plane hits from stereo superlayers appear to be offset from the reconstructed helix projection. The amount of the offset is proportional to the

Superlayer	Radius at center	# of cells	Stereo angle
1	46 cm	168	+3°
2	58 cm	192	0°
3	70 cm	240	−3°
4	82 cm	288	0°
5	94 cm	336	+3°
6	106 cm	384	0°
7	119 cm	432	−3°
8	131 cm	480	0°

Table 2.2: COT mechanical summary.

z coordinate. Proper matching of the stereo superlayer segments/hits onto existing $r-\phi$ track projection allows to extract z information of the track at 4 radial regions thereby measuring $\cot\theta$ and z_0 parameters of the helix.

A measure of the COT performance is the single hit position resolution. It has been measured to be about 140 μm, which translates into the transverse momentum resolution $\frac{\delta p_T}{p_T} \sim 0.15\% \frac{p_T}{[\mathrm{GeV/c}]}$.

It is crucial to mention that the global CDF coordinate system, introduced in the beginning of Section 2.2 is anchored to the center of the COT, *i.e.* the COT cylinder axis is the z axis of the global CDF coordinate system and the center of the COT is its origin. Positions of other detector components are measured (mechanically, by means of optical survey, or, when the ultimate precision is needed and the possibility exists, with data) with respect to the COT reference frame and encoded in so-called alignment tables. Whenever a spatial measurement is done by a system other than the COT it is usually done in the local reference frame and then converted into the global CDF (the COT) reference frame via the respective alignment table. One example of this procedure concerns the Silicon Vertex Detector, the other crucial tracking system, which we describe in the next section.

Silicon Vertex detector

For the purpose of a precise determination of the B hadron decay length one has to measure tracks in the vicinity of the beamline with excellent precision. Another important requirement is that a detector providing such measurement should be able to sustain large doses of radiation characteristic of the near-beamline region. Both of these criteria are satisfied by silicon micro-strip detectors which were pioneered in a hadron collider environment by CDF during Run I [79].

The principle on which "silicon" tracking is based is visualized in Figure 2-8. In a sense it is similar to that of the COT. The difference is that the position determination does not rely on the drift time measurement, but it rather relies on direct knowledge of the coordinates of the extremely small region where ionization is deposited.

In a typical silicon micro-strip detector finely spaced strips of strongly doped p-type silicon (p^+) are implanted on a lightly doped n-type silicon (n^-) substrate ($\sim$300 μm thick). On the opposite side a thin layer of strongly doped n-type silicon

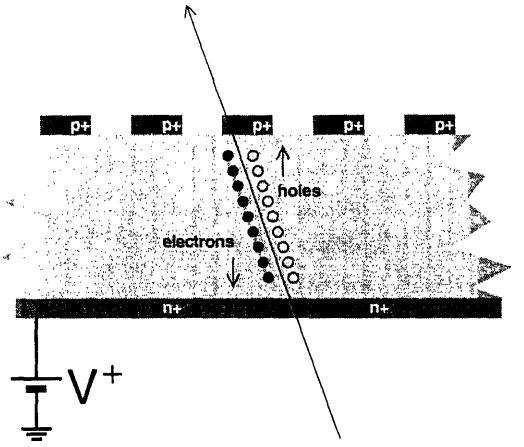


Figure 2-8: A generic silicon micro-strip detector.

(n^+) is deposited. A positive voltage applied to the n^+ side, which depletes the n^- volume of free electrons and creates an electric field. When a charged particle crosses the substrate it leaves a trail of electron-hole pairs from the ionization. In the electric field holes drift to the p^+ strips producing a well-localized signal. With additional effort the n^+ side can also be made in a form of orthogonal electron-collecting strips and put to use as a device measuring $r-z$ coordinate (assuming that p^+ side measures $r-\phi$).

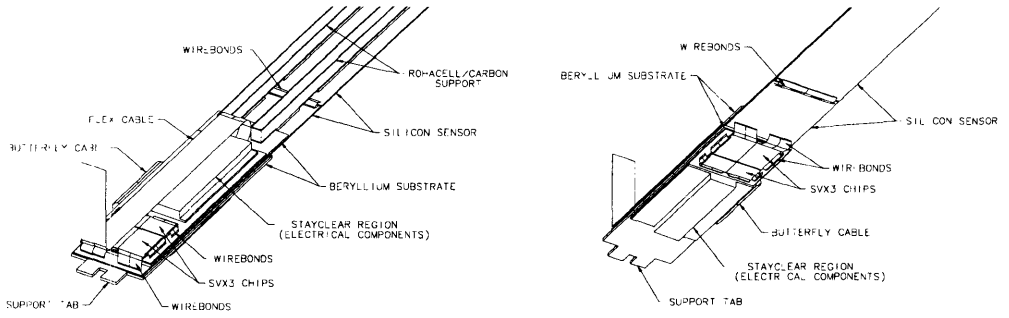
Usually the signal is detected on a small cluster of strips, rather than just one. The hit position is extracted by weighting the strip positions by the amount of charge collected. With this method the accuracy of the individual hit position measurement achieved at CDF is about 12 μm .

The CDF Silicon VerteX Detector, SVX [61], is built in three cylindrical barrels each 29 cm long. Each SVX barrel supports five layers of double-sided silicon micro-strip detectors between radii of 2.5 and 10.7 cm. For historical reasons layers are numbered from 0 (innermost) to 4 (outermost). Layers 0, 1 and 3 combine an $r-\phi$ measurement on one side with the 90° stereo ($r-z$) on the other, while layers 2 and 4 combine an $r-\phi$ measurement with small angle stereo at 1.2° (Table 2.3).

Layer	Radius, cm		# of strips		Strip pitch, μm		Stereo	Ladder active, mm	
	stereo	$r-\phi$	stereo	$r-\phi$	stereo	$r-\phi$	angle	width	length
0	2.55	3.00	256	256	60	141	90°	15.30	4×72.43
1	4.12	4.57	576	384	62	125.5	90°	23.75	4×72.43
2	6.52	7.02	640	640	60	60	+1.2°	38.34	4×72.38
3	8.22	8.72	512	768	60	141	90°	46.02	4×72.43
4	10.10	10.65	896	896	65	65	−1.2°	58.18	4×72.38

Table 2.3: SVX mechanical summary.

The actual silicon crystals (also called wafers or sensors) are supported by light-weight substrates (Rohacell foam) in assemblies called ladders (Figure 2-9). There



(a) top $(r-\phi)$ (b) bottom $(r-z)$

Figure 2-9: Perspective view of the top $(r-\phi)$ and bottom $(r-z)$ side of a Layer 0 ladder.

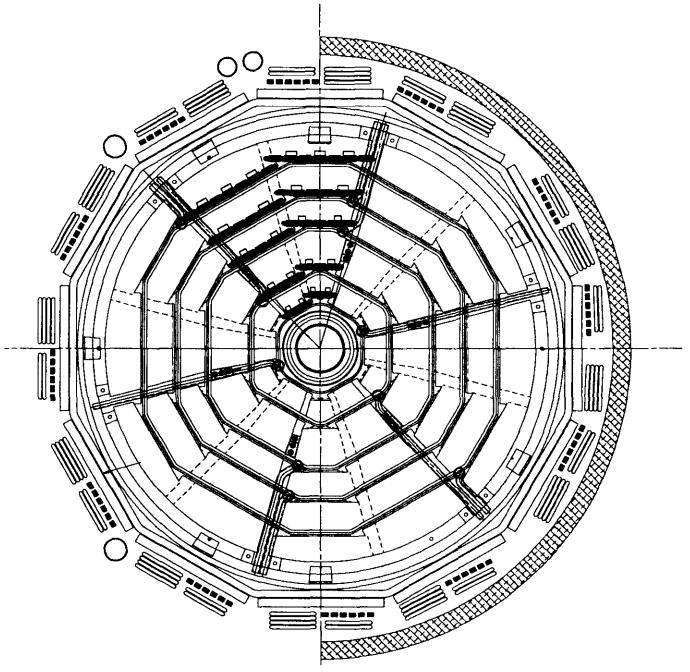


Figure 2-10: SVX bulkhead design. Placement of ladders is shown in two adjacent wedges.

are four wafers connected by the wirebonds in each ladder. Twelve ladders of the appropriate width comprise a layer. Sixty ladders are mounted between two precision-machined beryllium bulkheads (Figure 2-10) making an SVX barrel.

Each ladder is read out at both ends. The number of channels in the system (405,504) and the nature of the signals require that much of the electronics (charge-sensitive amplifiers, ADCs, dual-ported pipelines) is physically mounted on the system as opposed to somewhere outside the detector. It would simply be impossible to route that many cables to the outside of the detector, and signal would be lost in the noise

if not amplified beforehand. The negative side-effect is additional multiple Coulomb scattering caused by all the extra material in the tracking volume. Additionally, electronics dissipates a significant amount of heat, over 1 KW. To provide adequate cooling SVX design incorporates cooling channels into the bulkheads.

Every effort is made to place the SVX barrels in a coaxial manner. Remaining (small) spatial shifts are accounted for by barrel-to-barrel alignment. For the purpose of triggering[6] it is much more important to place SVX symmetrically around the beam than that of the z axis of the detector. This results in the fact that the SVX axis is offset from the COT axis by about the same amount that the beamline differs from the z axis of the detector.

When all SVX hits (some of which may be fakes) are found, the next step is to combine SVX and COT information in order to obtain the most complete helical fit for the track [80]. At this point it is absolutely important to have a very good alignment table, so that positions of the SVX hits originally measured only with respect to individual ladders are translated correctly into the global (COT) reference frame. The process starts by taking a track reconstructed by the COT alone, so-called COT-only track. This track (in fact, any track) is not just a line in space. Because all of the track parameters are measured with uncertainties (and respective correlations), the track is more like a tube of radius σ, determined by the errors on the track parameters. As additional complication, σ does not have to be (and in fact is not) the same in the ϕ and z directions.

The COT-only track is extrapolated through the SVX. As extrapolation proceeds from the outermost SVX layer toward the inside, the track error matrix is updated to reflect the amount of scattering material traversed. At each SVX layer hits that are within a certain radius (often defined as some multiple of σ rather than an absolute number because σ changes as hits are being added) are appended to the track and track re-fit is performed. A new track candidate is generated for each of the newly appended hits, but only the best two (in terms of the fit quality and the number of hits) are retained. Each of these candidates is then extrapolated further in, where the process is repeated. At the end there may still be several candidates associated to the original COT-only track. In this case the best one first in terms of the number of hits and then in terms of fit quality is selected.

A measure of the SVX performance is the impact parameter resolution. Presently it is at about $40\,\mu$m, which includes $25 - 30\,\mu$m contribution from the beam width. Another benchmark is the z_0 resolution[7], which is at the very respectable $70\,\mu$m.

Precise determination of track parameters allows one to discern which tracks came from what vertex and thereby distinguish the *primary* vertex (PV) from the *secondary* vertex (SV), such as a B hadron or other long-lived particle decay vertex.

[6] Certain triggers rely on the d_0 track parameter measured by the SVX. In case the beamline does not coincide with the SVX axis d_0 (and, consequently, the triggers) become ϕ-dependent!

[7] The quoted z_0 resolution has been obtained from track that have ISL hits on them. SVX stand-alone z_0 resolution is somewhat worse.

Intermediate Silicon Layer and Layer 00

Unlike the COT and the SVX, the Intermediate Silicon Layer (ISL) and Layer 00 (L00) do not have Run I predecessors. Layer 00 is also a "beyond baseline" project. As a consequence both systems have been subjects to various sorts of constraints: space, time, money, *etc.* On top of that there were some unforeseen complications with cooling for the ISL. Due to these circumstances ISL and L00 got integrated into detector operations relatively late. Still, because these systems broaden physics reach and potentially improve the quality of analyses that involve tracking, we present a brief description of them.

ISL In the central region a single ISL layer is placed at a radius of 22 cm. In the region $1.0 < |\eta| < 2.0$ two layers of silicon are placed at radii 20 cm and 28 cm, as indicated in Figure 2-4. Double sided silicon is used with 55 μm strip pitch on the axial side and 73 μm pitch on the stereo side with a 1.2° stereo angle. Only every other strip is read out to reduce the total channel count to 268,800. This affects the single hit resolution, which is about 16 μm on the axial side and about 23 μm on the stereo side, despite the fact that ISL ladders are similar to those of the SVX.

L00 The design of Layer 00 has six narrow (128 channel) and six wide (256 channel) groups in ϕ at radii 1.35 cm and 1.62 cm respectively (Figure 2-11). There are six readout modules in z, with two sensors bonded together in each module for a total length of 95 cm. The sensors are mounted on a carbon-fiber support structure which also provides cooling. L00 sensors are made of light-weight radiation-hard silicon (different from that of the SVX) and are single-sided with a 25(50) μm implant(readout) pitch.

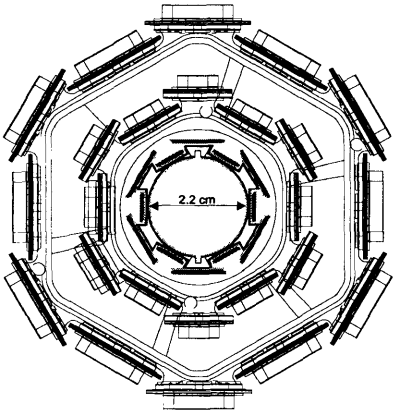


Figure 2-11: Transverse view of Layer 00 (innermost layer).

ISL improves tracking in the central region and in conjunction with the SVX allows stand-alone silicon tracking in the the region of $1.0 < |\eta| < 2.0$. L00 allows to overcome the effects of multiple scattering for tracks passing through high density regions of SVX (where the bulkheads and readout electronics is located) making possible d_0 resolutions as small as 25 μm .

2.2.2 Muon detection

Muons couple directly to the intermediate vector bosons W and Z. Being cleanly identifiable (as discussed below) they often serve as an indicator of an event involving $W \to \mu\bar{\nu}_\mu$ or $Z \to \mu^+\mu^-$ decays, allowing to trigger on such events. In the same way muons are a characteristic signature of the weak decays of heavy quarks, such as $t \to bW$ and $b \to cW$[8]. Triggers requiring two muons allow to collect a large statistics sample of $J/\psi \to \mu^+\mu^-$ decays which constitutes the basis for a large portion of the CDF B physics programme and is extensively used for detector understanding/calibration.

Muons are roughly 200 times heavier than electrons and therefore undergo ~40000 times less bremsstrahlung radiation. They are not subject to strong interaction with atomic nuclei either. Thus, muons can penetrate much more material than any other charged particles. This unique property of them is exploited in the CDF detector design, which places muon detectors behind the calorimetry[9] and, in some cases, additional steel absorbers.

CDF uses four systems of scintillators and proportional chambers in the detection of muons over the region of $|\eta| < 2$:

CMU – Central MUon detector,

CMP – Central Muon uPgrade,

CMX – Central Muon eXtension,

IMU – Intermediate MUon detector.

Design parameters of these systems are summarized in Table 2.4. Scintillators, placed next to the muon chambers, are used to suppress backgrounds coming from out-of-time interactions in the beam pipe material. The name of the partner scintillator system is obtained from the name of the muon system by replacing letter M (muon) with S (scintillator).

Figure 2-12 compares muon coverage in Run I with that of in Run II. Coverage in the central region ($|\eta| < 1$) is improved. IMU is totally new for Run II.

Individual muon candidates are detected in the muon chambers from the track segment they leave (typical di-muon event is shown in Figure 2-13). This track segment is called a *muon stub*. Having a stub does not guarantee a muon, because a stub can be due to a hadronic punch-through or just noise in the electronics. If a stub matches a certain track measured by the COT then the two are combined to make a muon. The details of this association are discussed in Section 2.3.1.

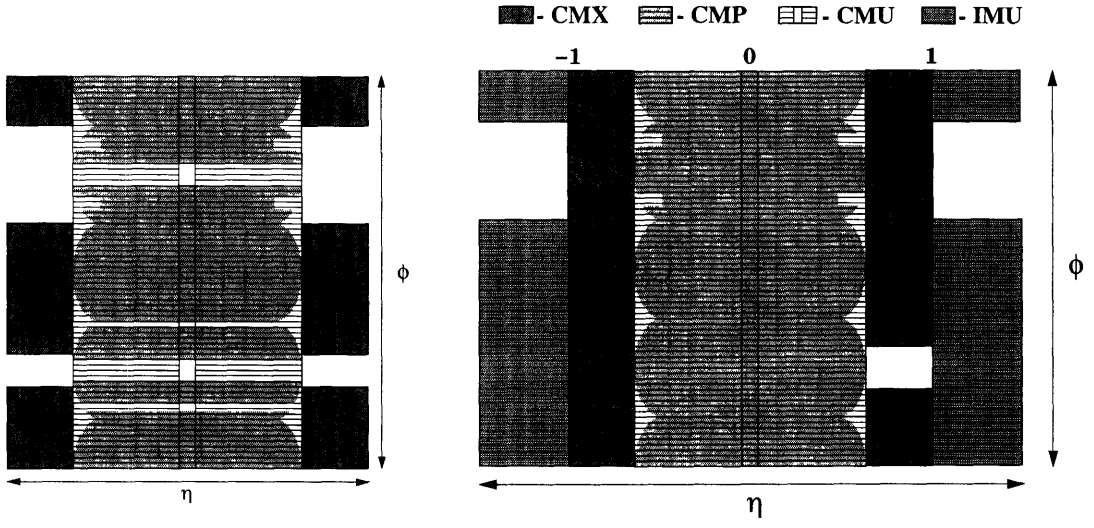
Although heavy shielding provides for "clean" muons, it introduces two issues of concern. First, it too causes muons to lose their energy (though at much slower rate than hadrons), so that muons below certain p_T threshold (different for different

[8] In the last case the W boson is virtual.

[9] Calorimeters usually work by fully absorbing particle's energy. In other words, most particles, but not muons, normally stop in the calorimeter.

	CMU	CMP/CSP	CMX/CSX	IMU
Pseudo-rapidity coverage	$\|\eta\| < 0.6$	$\|\eta\| < 0.6$	$0.6 < \|\eta\| < 1$	$1 < \|\eta\| < 1.5$
Drift tube cross-section	$2.7\times6.4\,\text{cm}^2$	$2.5\times15\,\text{cm}^2$	$2.5\times15\,\text{cm}^2$	$2.5\times8.4\,\text{cm}^2$
Drift tube length	226 cm	640 cm	180 cm	363 cm
Maximum drift time	800 ns	1.4 μs	1.4 μs	800 ns
# of drift tubes	2304	1076	2208	1728
Scint. counter thickness		2.5 cm	1.5 cm	2.5 cm
Scint. counter width		30 cm	30-40 cm	17 cm
Scint. counter length		320 cm	180 cm	180 cm
# of counters		269	324	864
Pion interaction length	5.5	7.8	6.2	6.2-20
Minimum muon p_T	1.4 GeV/c	2.2 GeV/c	1.4 GeV/c	1.4-2.0 GeV/c
Multiple scat. resolution	$12\,\text{cm}/p_T$	$15\,\text{cm}/p_T$	$13\,\text{cm}/p_T$	$13\text{-}25\,\text{cm}/p_T$

Table 2.4: Design parameters of the CDF muon systems. Pion interaction lengths and multiple scattering are quoted for a reference angle of $\theta = 90°$ in CMU and CMP/CSP, for an angle of $\theta = 55°$ in CMX/CSX and show the range of values for IMU.



(a) Run I

(b) Run II

Figure 2-12: Muon coverage in Run I (a) and in Run II (b).

muon systems, refer to Table 2.4), called *rangeout threshold*, never reach muon detectors. This is not a tremendous problem, because most of the muons of interest, the ones we trigger on, are required to have fairly high transverse momenta anyway. The other issue is multiple Coulomb scattering, the effect of which is amplified with introduction of additional material. Multiple Coulomb scattering randomly deflects

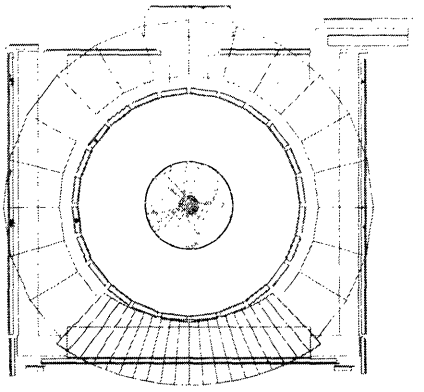


Figure 2-13: A typical event with a J/ψ candidate. Clusters of the purple dots outside the main tracking volume represent the hits registered in the CMU (inner ring) and CMP (thin outer square) muon systems.

a particle (muon in our case) traveling in the material, so that its point of entry into the muon system differs from what one would expect by naive extrapolation of the trajectory from the tracking volume. It complicates track-to-stub matching, but since the mismatch is roughly Gaussian and fairly narrow it can be taken care of.

The remainder of this section is devoted to a more detailed description of each of the muon systems.

CMU

The Central MUon (CMU) detector is located around the outside of the hadron calorimeter at a radial distance of 347 cm. Muon rangeout threshold for the amount of material provided mostly by CDF calorimetry is about 1.4 GeV/c.

The CMU is segmented in ϕ into 24 wedges, each covering 15°. The instrumented part of a wedge, as can be seen in Figure 2-14, is only 12.6°. This results in 24 ϕ gaps in the CMU coverage, 2.4° each. CMU is also divided into East (positive z) and West (negative z) halves with about 18 cm gap centered at $z = 0$ between the two.

Inside a wedge the detector is further segmented into three 4.2° chambers. Each of them consists of four layers of four rectangular drift cells as shown in Figure 2-15. CMU cells use the same gas mixture (Ar-Ethane) as the COT and they are run in proportional mode. The C/I-shaped aluminum cathodes at the edge of each cell are kept at -2500 V. The sense wire is made of stainless steel and is kept at +2325 V. The arrangement of the sense wires is such that two of them, from alternating layers, lie on a radial plane which passes through the z axis. The other two wires lie on the radial plane which is offset from the first one by 2 mm at the midpoint of the chamber. This allows to resolve the ambiguity as to which side of the sense wires a track passes by determining which two sense wires registered hits first.

Adjacent channels in the CMU are ganged together so that charge division can be

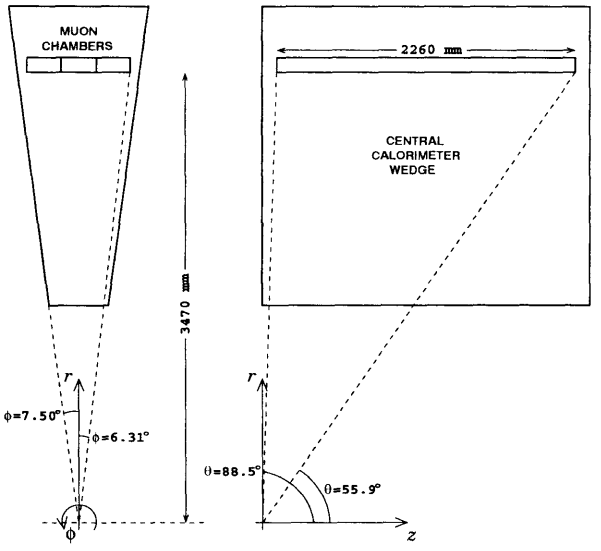


Figure 2-14: Geometry of a CMU wedge. $r-\phi$ and $r-z$ projections.

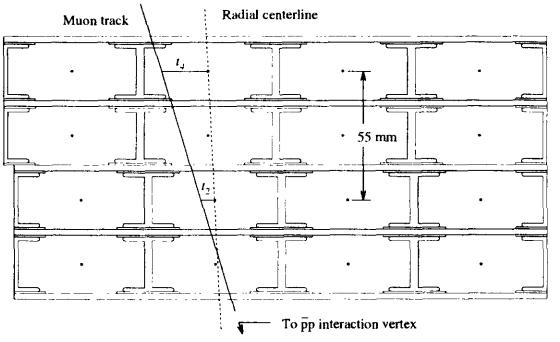


Figure 2-15: A single CMU chamber. This can be viewed as the zoom in onto one of the three small rectangular boxes in the $r-\phi$ projection of a wedge in Figure 2-14.

used to determine the z position of a hit on the wire. As additional benefit this allows to place readout electronics (amplifier, ADC and TDC) at one side of the chambers. Hit position resolution achieved by the CMU chambers is about $250\,\mu$m in the $r-\phi$ direction and of the order of 1 mm in z.

Further information about the CMU detector can be found in Reference [81].

CMP

As discussed in Section 2.2.2 and quoted in Table 2.4 CDF calorimetry provides "only" 5.5 hadron interaction lengths. To reduce contamination of high p_T muons from hadronic punch-throughs additional steel absorber followed by a layer of muon chambers has been implemented. This system has been called Central Muon uPgrade, CMP. For Run II CMP coverage has been expanded to close azimuthal gaps wherever

possible (compare (a) and (b) in Figure 2-12).

The CMP has a shape of a rectangular box with walls having equal length in z. The analogy with a box is only approximate (see Figure 2-13), for various reasons (location of cables, serviceability, *etc.*) some of the CMP walls are actually made of several sections which do not necessarily lie in the same plane.

At the top and bottom magnet return yokes serve as the absorbers. On the north and south (vertical) sides chambers are mounted on two movable non-magnetized steel walls. The additional steel increases absorption by 2.3 interaction lengths to a total of 7.8 interaction lengths while lifting the rangeout threshold to only 2.2 GeV/c.

CMP chambers are comprised of rectangular (2.5 cm×15 cm), single-wire drift tubes configured in four layers with alternate half-cell staggering. They are typically 640 cm long with few shorter sections on the bottom to avoid obstructions. The chambers are run in proportional mode and signals are read out by a single TDC per wire.

Because of the rectangular shape CMP pseudo-rapidity coverage varies as a function of ϕ (Figure 2-12). The important thing though is that CMU ϕ gaps are now covered, and large fraction of CMU muons with p_T above 2.2 GeV/c get "CMP confirmation," which allows for extremely clean muon selection.

Further information about the CMP detector can be found in Reference [82].

CMX

Central Muon eXtension, CMX, is another upgrade to the original detector designed to cover pseudo-rapidity range of $0.6 < |\eta| < 1$. The original CMX consisted of a pair of retractable arches at both East and West ends of the CDF. The CMX chambers lie on the surface of an imaginary cone whose apex is on the z axis of the detector at large positive(negative) z for the East(West) side. The cone's opening angle is about 45°.

In Run I the azimuthal coverage of CMX had a 30° gap at the top associated with necessity to route the Main Ring around (but in the vicinity of) the detector and the liquid helium lines (on the East side). There also was a 90° gap at the bottom of the detector where the conical sections are interrupted by the floor of the collision hall (Figure 2-3). For Run II these gaps are covered by the KeyStone (top-West) and the MiniSkirt (bottom-East, -West) additions (compare (a) and (b) in Figure 2-12). While KeyStone just integrates two missing CMX wedges, the MiniSkirt (bottom finer segmentation part in Figure 2-13) has a slightly different from the standard geometry.

Each 15° wedge (in the arches) is made up of eight layers of drift tubes with six tubes in each layer. The tubes have a rectangular cross-section of 2.5 cm×15 cm and are staggered so that at any azimuth coverage is at least four layers deep. The tubes used in CMX differ from those used in the CMP only in length; they are 180 cm long.

More details on the CMX detector can be found in Reference [82].

IMU

Introduction of the Intermediate Muon System, IMU, covering pseudo-rapidity region of $1 < |\eta| < 2$ with fine granularity is motivated by ability of the CDF tracking systems (in particular with the introduction of ISL) to reconstruct trajectories with $|\eta| > 1$ in the solenoid, thereby providing momentum measurement.

IMU consists of a barrel of drift chambers and scintillator counters around two steel toroids with additional counters between the toroids on either side of the CDF (see Figure 2-3). The IMU chambers and counters are virtually identical to those of the central muon systems and use the same electronics.

Specifics of the IMU design can be found in Reference [61].

2.2.3 Other systems

In this section we review the most significant of the remaining systems. Because these are not directly involved in the analysis presented, only a brief summary is given.

Calorimetry

CDF calorimetry has a uniform pattern of matched towers of EM and hadron calorimeters pointing back to the interaction region as can be seen in Figure 2-3. The calorimetry detectors are mechanically subdivided into three regions: central, wall and plug (in order of increasing pseudo-rapidity). They are called the Central Electro-Magnetic (CEM) and Central HAdron (CHA), Wall HAdron (WHA), Plug Electro-Magnetic (PEM) and Plug HAdron (PHA) calorimeters.

Central calorimeters cover 2π in azimuth and go as far as 1.1(1.3) in $|\eta|$ for EM(hadron) calorimeter. The tower size is $0.1 \times 15^\circ$ in $\eta \times \phi$. Plug calorimeters extend pseudo-rapidity coverage to $|\eta| = 3.6$ and use variable segmentation optimized for physics targeted.

CEM and PEM use lead sheets interspersed with scintillator as the active medium and employ phototube readout. Both calorimeters are equipped with shower maximum detectors, CES and PES for central and plug regions respectively, embedded at the depth of about six radiation lengths (X_0). Shower maximum detectors contribute to $e^{\pm}/\gamma$ identification, using the position measurement to match with tracks. They also provide transverse shower profile measurement to improve γ/π^0 separation and pulse height to help identify electromagnetic showers. CES (Central Electromagnetic Strip) is a gas multiwire proportional chamber with strip readout along the beamline and wire readout in the ϕ direction. The position resolution in both directions is about 2 mm. PES is made of scintillator strips read out by wavelength shifting fibers and features resolution near 1 mm.

There is a set of multiwire proportional chambers in front of the CEM. These are called Central PReshower (CPR) chambers, and they sample the electromagnetic showers that begin in the solenoid magnet material ($1.075X_0$) in front of them, providing greatly enhanced photon and soft electron identification. PEM does not have a dedicated preshower detector. Instead, its first layer of scintillator is made

thicker (10 mm) to yield more light and is read out separately. By analogy with CPR this system is called PPR; it provides for good efficiency in distinguishing between γ and π^0.

Basic quantities for EM calorimeters are summarized in Table 2.5.

	Thickness	# of layers	Resolution (E in GeV)
CEM	$19X_0$	20-30 Pb:3mm, Scint.:5mm	$13.5\%/\sqrt{E\sin\theta}\oplus 2\%$
PEM	$21X_0$	22 Pb:4.5mm, Scint.:4mm +Preshower	$16\%/\sqrt{E\sin\theta}\oplus 1\%$
CHA/WHA	$4.7/4.5\lambda_0$	32/15 Fe:25/50mm, Scint.:10mm	$75\%/\sqrt{E\sin\theta}\oplus 3\%$
PHA	$7\lambda_0$	23 Fe:51mm, Scint.:6mm	$80\%/\sqrt{E\sin\theta}\oplus 5\%$

Table 2.5: Basic quantities characterizing CDF II calorimetry.

The hadron calorimeters are located immediately behind the EM ones. They too are iron/scintillator sampling devices. Major characteristics of CDF hadron calorimeters are included in Table 2.5.

Useful literature on calorimetry and its applications at CDF includes References [61, 83].

TOF

The Time-of-Flight system, TOF, expands CDF's particle identification capability in the low p_T region. TOF measures arrival time t of a particle with respect to the collision time t_0. The mass m of the particle is then determined using the path length L and momentum p measured by the tracking system via relationship

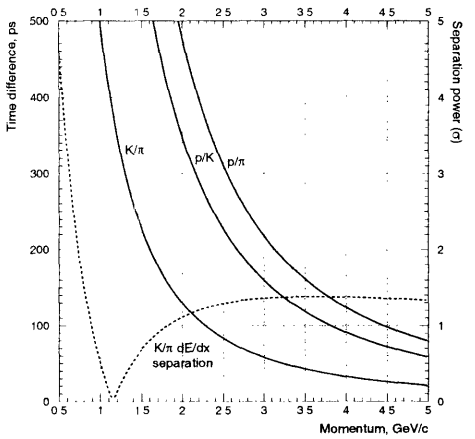
$$m = \frac{p}{c}\sqrt{\frac{(ct)^2}{L^2} - 1}. \tag{2.9}$$

The TOF consists of 216 BC-408 scintillator bars installed at a radius of about 138 cm (from the z axis) in the 4.7 cm space between the outer shell of the COT and the cryostat of the superconducting solenoid (Figure 2-4). Bars are approximately 279 cm long and $4 \times 4\,\mathrm{cm}^2$ in cross-section. With cylindrical geometry TOF provides 2π coverage in ϕ and roughly $(-1; 1)$ coverage in η.

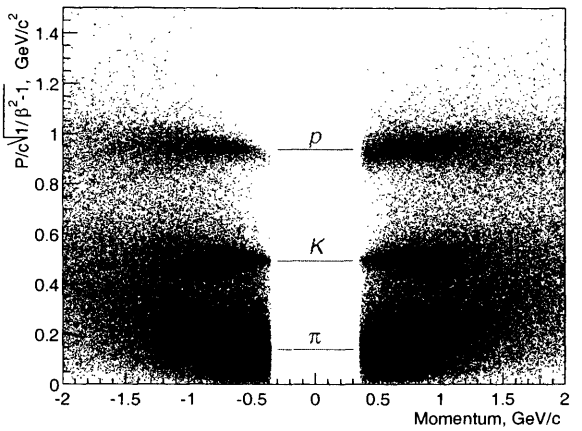
Bars are read out at both ends by fine-mesh PMTs (Hamamatsu R7761), capable of providing adequate gain even if used inside 1.4 T magnetic field. Usage of long attenuation length fast rise time scintillator in conjunction with these PMTs allowed to achieve specified resolution of 100 ps.

Figure 2-16 (a) shows time-of-flight difference for K/π, p/π and K/p hypotheses and the separation power assuming resolution of 100 ps. In Figure 2-16 (b) early TOF performance is illustrated.

More details on TOF construction and performance can be found in Reference [84].



(a) K/π, p/π and K/p time difference over a path of 140 cm as a function of momentum, expressed in ps and separation power (σ), assuming resolution of 100 ps. The dashed line shows the K/π separation power from dE/dx measurement by the COT.



(b) TOF reconstructed mass vs. momentum for positive and negative tracks. Clusters of data points corresponding to π, K and p are clearly visible. Data are from Tevatron store 860 (12/23/2001).

Figure 2-16: Time-of-Flight system performance: design (a) and data (b).

CLC

The Cherenkov Luminosity Counter, CLC, measures the average number of interactions per bunch crossing μ. Then instantaneous luminosity $\mathcal{L}$ is extracted using formula

$$\mu \cdot f_{bc} = \sigma_{p\bar{p}} \cdot \mathcal{L}, \tag{2.10}$$

where $\sigma_{p\bar{p}}$ is relatively well known total $p\bar{p}$ cross-section at $\sqrt{s} = 1.96\,\text{TeV}$ and f_{bc} is the rate of bunch crossings in the Tevatron.

The CLC utilizes the effect, known as Cherenkov radiation — particles traversing a medium at a speed higher than the speed of light in this medium radiate light into a cone around particle direction; cone's opening angle depends on the ratio of the two speeds and the refraction index of the medium. The idea is to use an assembly of long gas Cherenkov counters positioned in the Plug Calorimeter 3° gap so that they point toward the interaction region as schematically shown in Figure 2-17. This arrangement allows to make the detector much more sensitive to the particles coming directly from the interaction point because their flight path in the gas of the counter is the longest and therefore the amount of light produced — the largest.

Excellent timing resolution (~50 ps) and clever design allow the CLC to discern multiple interactions within the same bunch crossing and achieve an overall accuracy of the luminosity measurement better than 5%.

In depth information on the CLC design and performance is given in Reference [85].

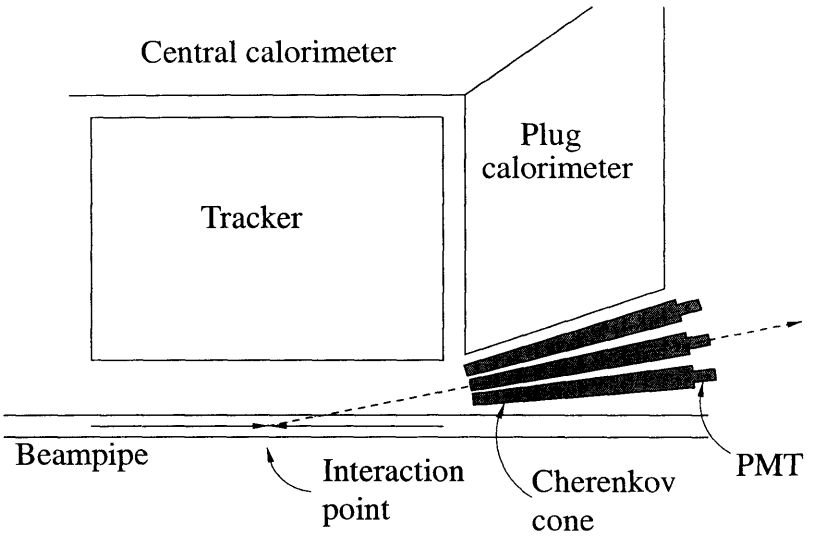


Figure 2-17: CLC geometry.

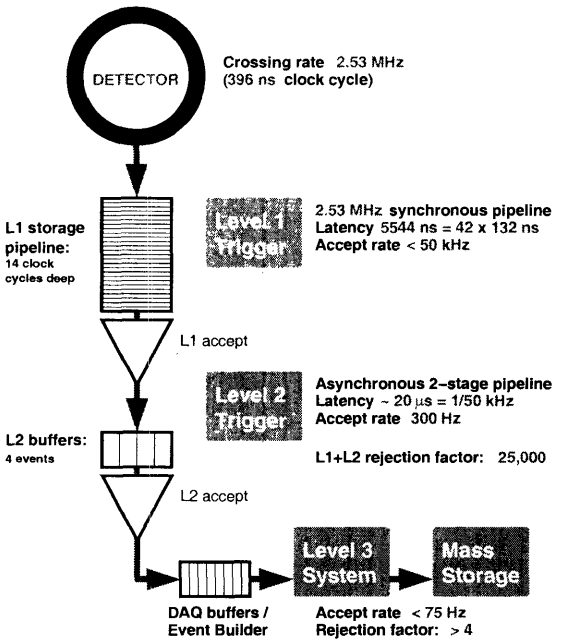
2.3 Triggers and data acquisition

In the Tevatron proton and antiproton bunches cross every 396 ns. Using $\sigma_{p\bar{p}}$ of the order of 100 mb (at $\sqrt{s} = 1.96$ TeV) in the Equation 2.10 one finds that the instantaneous luminosity of $(2-3)\times10^{31}\text{cm}^{-2}\text{s}^{-1}$ corresponds to about one interaction per crossing or one event every 396 ns. If the Tevatron switches to the mode of operation with 132 ns bunch spacing the number of interactions per unit time is going to increase, but even now $p\bar{p}$ collisions produce an enormous data rate.

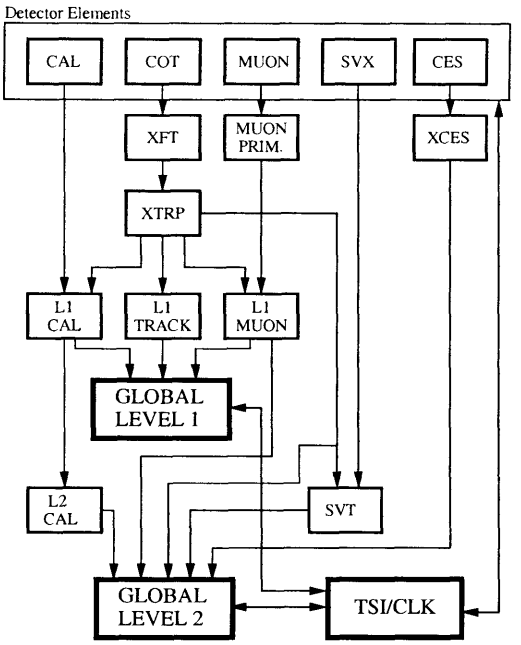
Unfortunately, by far not all events can be recorded. One reason is that this is just beyond the current mass storage and throughput capabilities. Indeed, with the average size of an event of the order of 100 KB the required bandwidth would be 250 MB/s, which is not only difficult and expensive to handle by itself, but it would also result in 150 TB of data in one week of running! The second impediment is that it takes about 2 ms to read the entire detector out, a time long enough for another 5,000 or so interactions to happen (and go unrecorded).

However, most processes of interest have cross-sections in the range of $10-100\,\mu$b or smaller, *i.e.* at least three orders of magnitude smaller than the total $p\bar{p}$ cross-section. This fact makes it possible to circumvent the above two problems by reading out and storing only the interesting events making on-the-fly selection. Of course, it is not feasible to make a good judgment whether an event is interesting or not in 396 ns before the next one comes. Thus the design of the fully pipelined data acquisition system and three-level deadtimeless trigger pictured in Figure 2-18 (a).

Level-1 trigger discards the vast majority of the events, those which do not have any interesting signatures. Since it can take as long as $5.5\,\mu$s for the Level-1 to reach its decision all front-end electronics are equipped with buffers 14 (396 ns) bunch crossings deep. If the Level-1 does not reject an event it proceeds to the Level-2. The Level-2 does a more careful analysis of the information taking longer time (few tens of microseconds) to do so. Only if its criteria are met is the entire detector read out.



(a) CDF readout functional block diagram.



(b) Block diagram of the CDF trigger system.

Figure 2-18: Block diagram detailing CDF DAQ and L1/L2 trigger system.

The Level-1/Level-2 trigger system is managed by the Trigger Supervisor Interface (TSI/CLK box in Figure 2-18 (b)). The TSI also provides an interface between the triggers and the DAQ as well as global clock and bunch crossing signal.

In the case of Level-2 accept the digitized output from all detector components is collected and transferred to the Event Builder. There the event fragments obtained from different subsystems are organized into a properly ordered sequence, which is then fed to the Level-3 computing farm. At the Level-3 the event undergoes fairly thorough reconstruction after which the final requirements are applied. If the latter are satisfied the event gets written to a mass storage device.

Below we discuss CDF trigger and data acquisition system, especially parts relevant to our analyses, in greater details. The conceptual design of the entire system is presented in Reference [87] and an elaborate description is given in Reference [61].

2.3.1 Level-1

The Level-1 is a synchronous system with an event read in and an accept/reject decision made (on an earlier event) every bunch crossing. The decision is based on transverse energy in the calorimeters, tracks in the COT and stubs in the CMU, CMX and CMP chambers. Level-1 uses rudimentary (no detailed reconstruction) versions of the above mentioned objects, collectively called *primitives*.

In several cases track primitives are combined with calorimeter or muon primitives

to form e, μ and jet objects as shown by arrows in Figure 2-18 (b). Given the time constraints the ability to combine information from different sub-detectors is rather limited, but this is an important step in the rate reduction.

To ensure the maximum speed Level-1 uses custom designed hardware. The total rate of Level-1 accepts is about few tens of KHz, a factor of few hundred smaller than the input rate of 2.5 MHz.

There are many algorithms (or, equivalently, individual triggers) at the Level-1: one looking for two high-p_T tracks, one looking for a single high-energy muon, one looking for large missing E_T, *etc.* [86]. The one used in the analyses of this dissertation is a *di-muon*, *i.e.* requiring two muons, trigger. We discuss it more specifically below, right after the introduction of the two systems, XFT and XTRP, which provide information to be incorporated into decision making process for several separate Level-1 triggers, including the di-muon trigger.

XFT

EXtremely Fast Tracker, the XFT, identifies track primitives – high p_T tracks ($p_T \approx$ 1.5 GeV/c and above) in the $r-\phi$ view using the four axial superlayers of the COT. Track identification is accomplished in two steps: segment finding and segment linking. In the finding step all COT axial hits are classified as either "prompt" (drift time $<$ 44 ns) or "delayed" (44 ns $<$ drift time $<$ 132 ns). Then a set of binary masks (predefined patterns of prompt/delayed hits) is applied in order to find all segments compatible with a valid track[10] passing through a given superlayer. For a successful match a "pixel" is set. This pixel represents the ϕ position of the segment plus slope information in the two outermost axial superlayers (needed to resolve the charge). In the second (linking) step four pixels appearing to have come from the same track are identified and a crude estimate of track parameters is done. Thus found tracks are reported to the XTRP unit and a copy of them is preserved to be used in the Level-2.

Further information about the XFT can be found in Reference [88].

XTRP

EXTRaPolator unit, the XTRP, is responsible for quick extrapolating of the XFT tracks to other detector subsystems such as calorimetry and muon chambers (Figure 2-18 (b)). Matching these tracks with electron towers (EM calorimeter) or muon stubs (muon chambers) allows for somewhat more complex primitives and, therefore, an additional rejection at the Level-1. A detailed example of the XTRP operation is given in "Di-muon trigger" section which follows.

Di-muon trigger

Di-muon trigger at the Level-1 is actually a combination of the CMU-CMU (both muons from the CMU) and the CMU-CMX (one muon from the CMU and the other

[10]The one that originates in the vicinity of the beamline and has high enough p_T .

one from the CMX) triggers. Since the logic is very similar between these two we only explain the CMU-CMU trigger. Before we proceed it is useful to introduce some terminology pertaining to triggering on CMU muons:

Stack – an aggregate of four cells stacked on top of each other (Figure 2-15). There are 288 stacks on each (East and West) side of the detector.

Level-1 stub – a track segment in a stack such that at least cells 1 and 3 or cells 2 and 4 have hits separated in time by no more than the stub gate width of 396 ns.

Tower – a logical unit that corresponds to a pair of neighboring stacks. There are six towers in a wedge. The tower is said to have fired if at least one of the stacks it is comprised of has a Level-1 stub. Otherwise it is said to be empty.

Muon tower – a fired tower matched with an XFT track as reported by the XTRP unit (explained below).

There is no complicated stub reconstruction done at the Level-1; we only know which stacks have a stub (as opposed to later stages where stub position and slope with respect to the radial direction are measured). In fact, even this information is generalized and the trigger decision is based on which towers have fired. The CMU-CMU trigger ("L1_TWO_CMU_PT1.5") logic is as follows:

1. The XFT reports p_T, ϕ_{SL6} (ϕ at the COT superlayer 6) and charge of the tracks it has identified to the XTRP. The XTRP uses this information and the approximation that a track originated at the beamline to extrapolate the track to the inner radius of the CMU (347.8 cm). For each track the XTRP determines the "footprint" – a ϕ window in which the track could end up with 99.5%(3σ) probability due to the multiple scattering.

2. If at least one footprint found in step 1 overlaps with a fired CMU tower that tower is promoted into a muon tower.

3. If the event has at least two non-adjacent muon towers it gets accepted. "Non-adjacent" means the two are either separated by at least two other towers (which are not required to be empty) or they are on different sides of the detector. In this consideration a 2.4° gap between nearby CMU wedges is treated as a tower.

The CMU-CMX ("L1_CMU1.5_PT1.5_&_CMX1.5_PT2_CSX") trigger logic is similar, except that for the purpose of finding CMX trigger muon only XFT tracks with p_T greater than 2.2 GeV/c are considered; muons with smaller p_T range out as discussed in Section 2.2.2. The only other difference is that no azimuthal separation is required between the CMU and CMX muons.

More details on the design and the evolution of the Level-1 di-muon trigger can be found in Reference [89].

2.3.2 Level-2

The Level-2 is an asynchronous system which processes events accepted by the Level-1 in the time-ordered fashion. It uses Level-1 primitives as well as additional data coming from the shower maximum strip chambers (CES) in the central calorimeter and the $r-\phi$ strips of the SVX as indicated in Figure 2-18 (b). There are three hardware subsystems building primitives at the Level-2: L2CAL, XSEC and SVT (see Figure 2-18 (b)).

The L2CAL hardware receives trigger towers from L1CAL and finds energy tower clusters by applying seed and shoulder thresholds.

The XSEC system generates bitmap of strips above certain threshold in the shower maximum chambers (8 bits per 15°). Then XFT track primitives are extrapolated by the XTRP to the CES radius and fed to the XSEC where they are matched with the strip bitmap, producing electron candidates.

The SVT uses SVX $r-\phi$ hits to extend XFT track primitives inside the SVX volume, closer to the beamline. The SVT improves the XFT ϕ_0 and p_T resolutions and adds the measurement of the track impact parameter d_0 (original XFT track primitives are beamline constrained). Significant impact parameter indicates a displaced vertex, which is an extremely powerful cut/signature.

Primitive building hardware constitutes the first stage of the Level-2. The second stage consists of four programmable processors (Alpha), which operate on lists of objects provided by the first stage.

The Level-2 latency is $20-30\,\mu$s, which along with the buffers for four events allows to bring the event rate from tens of KHz down to $200-300$ Hz.

2.3.3 Event Builder and Level-3 trigger

In the event of a Level-2 accept the entire detector is read out thereby emptying a slot in all of the detector buffers for the next event. The just read out event fragments are put in the proper order by the ATM switch based system known as the Event Builder [90]. Thus arranged event fragments are channeled to the Level-3.

Level-3 [90] is a farm of conventional PCs running Linux. The farm is split in a number of (16 as of this writing) sub-farms of identical architecture. Each sub-farm consists of a head node (also known as a converter node) and $12-16$ processor nodes. There are also eight so-called output nodes in the Level-3, each is shared by two sub-farms.

It is the converter node that receives an ordered sequence of event fragments from the Event Builder. In the converter node these fragments are assembled in a block of data, called the *event record*, suitable for analysis by CDF software. From then on the event record becomes one and the only piece of information about a particular event.

Each event record is fed to one of the processor nodes in the sub-farm, where event reconstruction is performed and final trigger requirements are applied. At this stage Level-3 takes advantage of the full detector information and improved resolution not available at the lower trigger levels. If an event satisfies Level-3 requirements,

the corresponding event record gets passed to the output node which subsequently transfers it to the mass storage.

Parallel processing of many events by many nodes allows for much more time to look at a particular event and therefore for a much more accurate decision whether to keep it or not. Level-3 provides an additional reduction of about factor of 4 thus bringing the total event rate down to approximately 75 Hz.

It should be noted that detector is taking data for (usually large, but never exceeding duration of the store) continuous periods of time called *runs*. During the run detector configuration (including all the online calibration constants, *etc.*) is stable, it can change however from one run to another, affecting or not affecting the data being recorded. In any case, the events that pass Level-3 are being stamped with run number, so that at a later time information about detector configuration at the time the event was taken could be retrieved from the database.

2.4 Detector operation and off-line data handling

The data flow from the Level-3 is segmented into ten streams, which are denoted with letters from A to J. The events which passed all three levels of di-muon requirements are written out as "Stream J". All the streams are being written to tape in real time, as the data are being taken, *i.e.* in on-line regime.

All other manipulations with data are referred to as off-line data handling. The most important of these operations is *production*, which stands for processing all of the data with `ProductionExe` executable. At this stage the raw data banks are unpacked and (collections of) physics objects suitable for analysis, such as tracks, vertices, muons, electrons, jets, *etc.* are generated. This is similar to what is being done at Level-3, except it is done in a much more elaborate fashion, applying the most up-to-date calibrations, using the best measured beamlines, *etc.*

The output of the `ProductionExe` undergoes further categorization, such that ten data streams are splitted into 35 (as of this writing) data sets. Normally, it is a data set what an average collaborator deals with when (s)he performs a physics analysis. In this dissertation we consider the `jpmm0c` data set, which includes all events from Stream J. In practice it is more convenient to work with the compressed version of this data set called `xpmm0c`. The compression is achieved by means of dropping objects not used in most analyses from the event record. This greatly reduces the storage required (in particular allows disk storage, rather than tape storage) and speeds up the analysis, of which reading the data in is often the most or second most time-consuming part.

Occasionally, but not too often due to ever growing volumes of data, it is found beneficial to re-process part or all of the data with a newer and better version of `ProductionExe`. So one needs to state the production version to fully specify the data used in the analysis. We use `xpmm0c` data set produced with 5.1.0 version of `ProductionExe`.

Chapter 3

Analysis of the data

In order to conduct the desired studies of B mesons we need to isolate the events of interest, *i.e. signal*, from the multitude of other events occurring in the detector, commonly referred to as *background*. Unfortunately, such isolation cannot be perfect and a certain amount of background makes its way into the sample. The primary implication is that we have to find an appropriate description (in terms of mass, proper decay length and angular variables) of the background, which is necessary to prevent the background from affecting the extraction of the signal properties.

Having accepted the inevitable presence of the background in the sample, one needs to choose optimal selection, usually in terms of boundaries, also known as *cuts*, on certain quantities characterizing the events, such that the uncertainty on the extracted parameters of interest is minimized. Often making optimal selection does not translate into minimizing the amount of background, because of the two facts: first, once there is some background it may be beneficial to have a substantial amount of it such that its properties can be adequately modeled; second, using cuts to reduce the amount of background also reduces the amount of signal, such that the overall statistical power is diminished.

Another general analysis issue is related to the fact that all events (signal and background) are observed through a detector. A relevant analogy here would be looking at some small extra-terrestrial object of unknown nature with a lens that has spherical, chromatic and all other known aberrations plus, perhaps, some unknown ones instead of ideal aberration-free lens. This is to say that not only detectors have a finite resolution, but also they give a distorted view of events they register. One of the major goals of the analysis is to unfold these distortions and/or minimize them. The remaining distortions need to be quantified and included into the uncertainty on the measured quantities.

Of course, there are more issues in an analysis than just the above two. The current chapter covers the issues, arising in the analyses of this dissertation, in as much detail as reasonably possible. Some of the discussions are self-contained and, at times, highly technical. Wherever possible we move them over to one of the Appendices to preserve the flow of information and improve readability. A fair warning: certain issues are so intertwined that it is inevitable to have a few forward references.

Throughout this chapter, whenever a particular decay/particle is referred to, the

charge conjugated decay/particle is also implied, unless specifically stated otherwise.

3.1 Data and candidate selection

The three particular decay channels we are looking at are $B_u \to J/\psi K^+$, $B_d \to J/\psi K^{*0}$, and $B_s \to J/\psi\phi$. All three are reconstructed from the `xpmm0c` data set explained in Section 2.4. This analysis is limited to the inclusive run range 138809 – 179056. These runs were recorded in the period from February 9, 2002 to February 13, 2004. Only "good runs" have been selected. A run is considered to be good for a particular analysis if all relevant detector systems operated normally during this run as determined by the experts. The particular good run definition used in the current analysis, although not extremely meaningful for people outside the CDF Collaboration, is given in Appendix B. The B samples considered in this analysis correspond to the total integrated luminosity of $258 \pm 15\,\mathrm{pb}^{-1}$.

In $p\bar{p}$ collisions at $\sqrt{s} = 1.96\,\mathrm{TeV}$ with the trigger requirements we have, at least 80% of the J/ψ are from prompt $c\bar{c}$ production, while not more than 20% are coming from B decays [92]. This is the main reason why the `xpmm0c` data set, though compressed, is still rather bulky to deal with on a daily basis. To overcome this hurdle we have created skimmed samples (or, simply, skims) of B_u, B_d, and B_s candidates. In preparing a skim the selection code is run without the most detailed calibrations, but with loose requirements, such that nothing that can potentially pass the subsequent strict analysis cuts is excluded. In preparing the B_u, B_d, and B_s skims we use the following requirements:

- $p_T^{B_{u,d,s}} > 4.45\,\mathrm{GeV/c}$,
- $p_T^{K^+,K^{*0},\phi} > 1.45\,\mathrm{GeV/c}$,
- $p_T^{track} > 0.4\,\mathrm{GeV/c}$,
- wide mass windows on $B_{u,d,s}$ and J/ψ, K^{*0}, ϕ,
- $\geqslant 3$ SVX $r-\phi$ hits on each track comprising a B candidate.

After the skims are prepared the most complete and accurate version of the selection code is run. In fact we need to run it multiple times for various investigations and cross-checks, which justifies the existence of skims. This final version of the analysis code uses the best available calibrations. It also has the most detailed and accurate (also time-consuming) procedures, such as those described in Section 3.1.1, turned on. At the output we get a collection of B meson candidates, stored in a form of `ROOT` [93] `NTuple`. The `NTuple` is a data structure which in our case holds all the information of interest about each B meson candidate and allows efficient access to it. It is on the `NTuple` the final analysis is done.

3.1.1 Treatment of tracks

All tracks considered in the analysis are refit using `TrackRefitter` [94]. A number of goals are accomplished during this step:

1. Track parameters are corrected for the energy loss in the material of the detector using appropriate particle type hypothesis. Indeed, as a particle travels through the material of the detector it looses energy, therefore a measurement by a measuring station further downstream may need to be corrected before being used.

2. In tracking, the uncertainties on the COT hit positions are ascribed without taking into account the effect of multiple scattering in the material of the detector, therefore the uncertainties on track parameters returned by the track fit tend to be underestimated (in a fashion that depends where the track goes in the detector and how energetic the track is). `TrackRefitter` brings these uncertainties back in line using empirical corrections.

3. L00 and ISL hits are dropped in the track refit, because the respective tracking sub-systems are not yet sufficiently well aligned and calibrated. Also the results of the kind of analyses we describe are not expected to benefit substantially from using L00 and/or ISL.

For the lifetime measurement it is important to have good vertex resolution, therefore using only well-measured tracks is imperative. Precise and uniform in quality vertices are obtained by requiring each of the tracks making up a B meson candidate to have $r-\phi$ hits in at least three distinct SVX layers. The starting point here, however, is the fact that a well measured COT track (which one needs as a seed in order to find relevant SVX hits and make a good combined track) should not leave COT until after it crosses superlayer (SL) 6.

Muons are subject to even more stringent requirement because we trigger on them, which implies that XFT finds 11-12 hit segments in each of the SL2, SL4, SL6, and SL8. The CDF Tracking Group has established [95] that real, straight tracks have good segments in at least three axial SLs and at least two stereo SLs. The Tracking Group defines a good segment as the one having five or more hits. Such a small number is justified because "bad" hits are being dropped "aggressively". We have investigated the SL usage by XFT triggered tracks, such as muons from J/ψ (Figure 3-1) as well as "unbiased" tracks, such as kaons from ϕ (Figure 3-2). The Tracking Group recommendation to require at least three axial SLs with five or more hits as well as at least two stereo SLs with five or more hits on every track does make sense for muons if one wants near perfect efficiency with reasonable purity. However, for kaons and pions, which might exit the COT before reaching SL7, it seems reasonable to require only two SLs of each kind with at least five hits in each to preserve maximum efficiency.

The summary of the track quality requirements is included in Table 3.1.

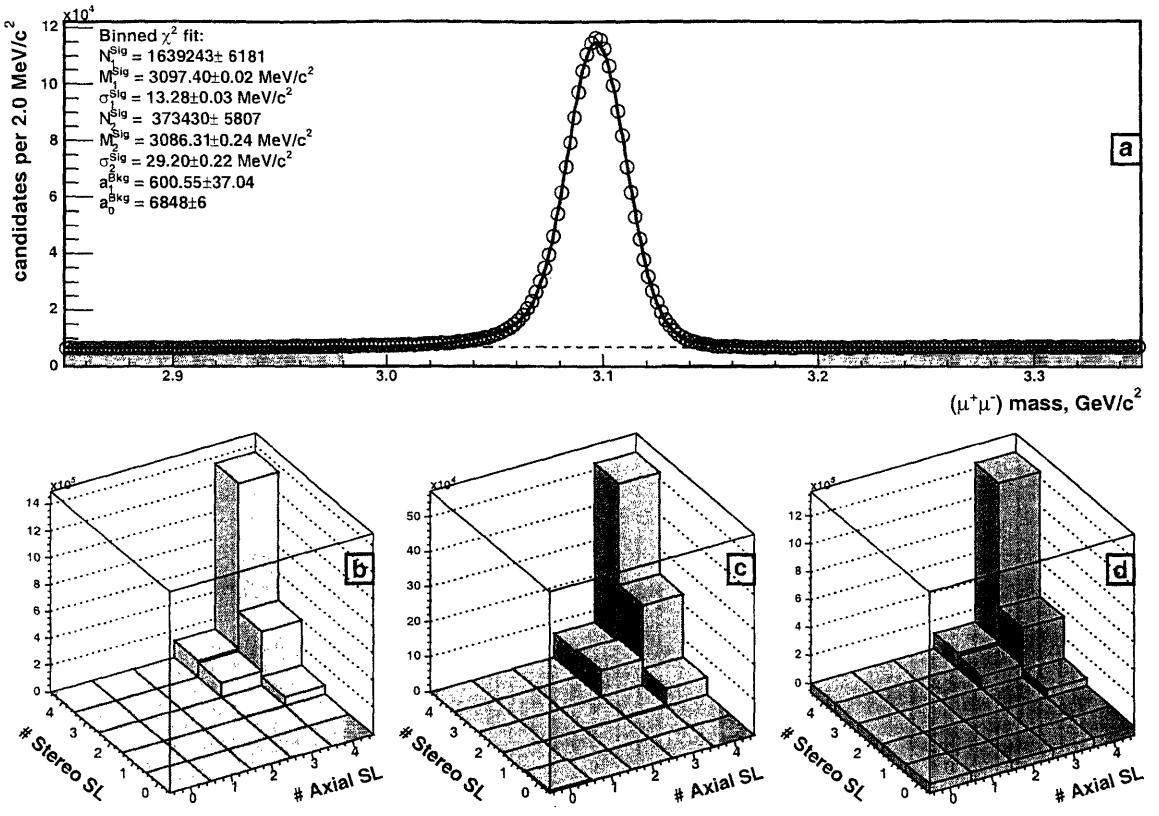


Figure 3-1: SL usage by muons from J/ψ. (a) Mass distribution of the J/ψ candidates used in the side-band subtraction. Five hits per SL is required. SL usage for: (a) signal region, (b) side-bands, (c) side-band subtracted signal.

3.1.2 Treatment of muons

On top of the nominal track quality cuts muons are subject to additional requirements. In particular, among CMU muon candidates only those are selected for which the track and the muon chamber stub are compatible within uncertainties. This condition is imposed by requiring that the χ^2 variable characterizing the match between the track and the stub is less than 9.0. Also, in addition to the tight COT hit requirement discussed in Section 3.1.1, the offline p_T of all muons is required to be greater than 1.5 GeV/c to reflect the fact that muons from J/ψ should have been found by the XFT and should not have ranged out. These requirements are collected in Table 3.1.

3.1.3 J/ψ, K^{*0} and ϕ selection

To suppress combinatorial background it makes sense to select only those intermediate daughter meson candidates that fall into a certain mass window. Such mass windows are set to $M_{PDG}^{J/\psi} \pm 80\,\mathrm{MeV/c^2}$ for a J/ψ candidate, $M_{PDG}^{K^{*0}} \pm 50\,\mathrm{MeV/c^2}$ and $M_{PDG}^{\phi} \pm 6.5\,\mathrm{MeV/c^2}$ respectively for K^{*0} and ϕ candidates. The J/ψ mass window is chosen

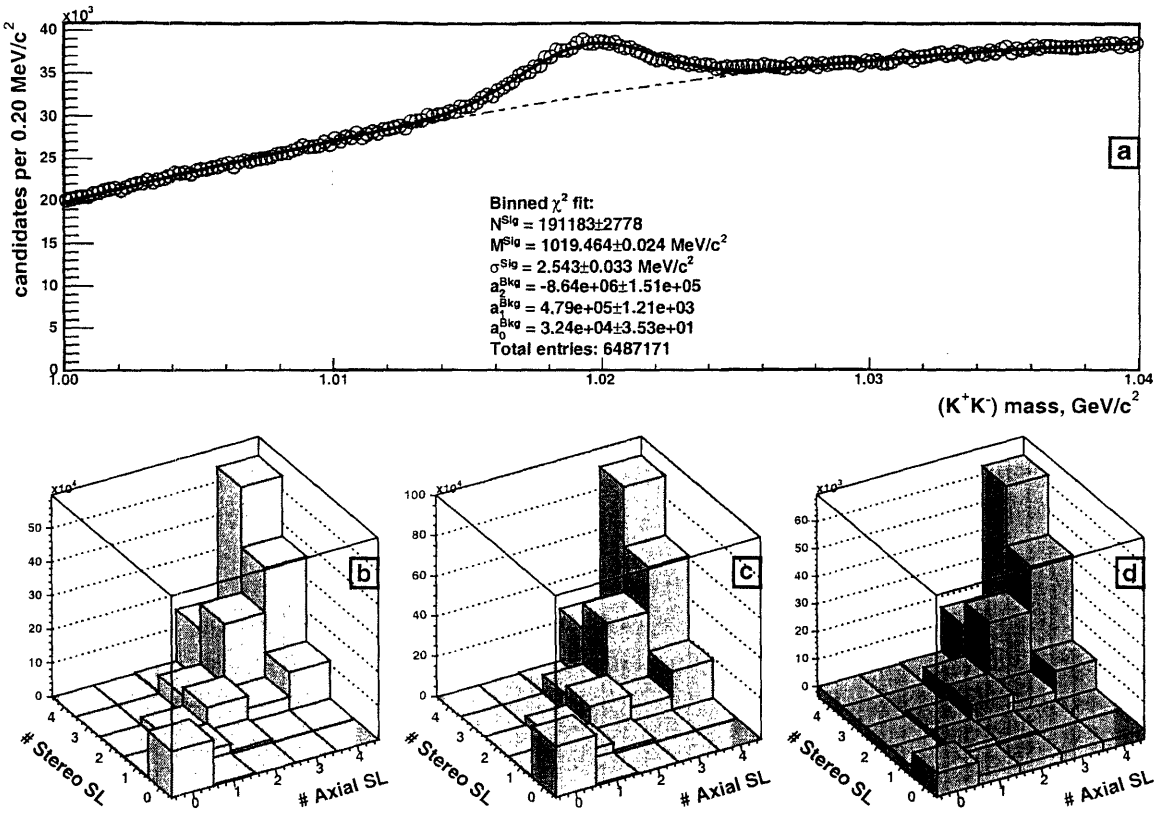


Figure 3-2: SL usage by kaons from ϕ. (a) Mass distribution of the ϕ candidates used in the side-band subtraction. Five hits per SL is required. SL usage for: (a) signal region, (b) side-bands, (c) side-band subtracted signal.

such that the highest p_T J/ψ candidates[1] as well as candidates from the radiative tail are still included into consideration. The K^{*0} and ϕ mass windows are motivated in Section 3.1.5 and explained in detail in Appendix E.

Particle identification (PID) capabilities of the CDF detector are rather limited. Not knowing if a given track is left by a (charged) π, K or p often leads to the necessity of trying more than one mass assignment, which in turn leads to larger backgrounds and other challenges.

A particular issue and a substantial complication for the current analysis related to the lack of PID arises in $K^{*0} \to K^+\pi^-$ reconstruction. The K^{*0} is fairly wide and it may happen that both $(K\pi)$ and (πK) mass assignments result in masses that fall into the specified mass window. In this case only the candidate with the mass closest to $M_{PDG}^{K^{*0}}$ is retained. The choice of this procedure has implications that are discussed in Section 3.2.4, Appendix I and various other places throughout this chapter.

[1]These have the worst mass resolution, but at the same time these are the likeliest to have originated from B decays [92].

3.1.4 B meson selection

The three or four daughter tracks (μ^+, μ^-, K^+ for B_u; μ^+, μ^-, K^+, π^- for B_d; and μ^+, μ^-, K^+, K^- for B_s) after passing technical quality requirements described above are combined in a B meson vertex fit, as described in Section 2.2.1. In this fit the di-muon invariant mass is constrained to $M_{PDG}^{J/\psi}$ in order to improve the B meson mass resolution. If the vertex fit is successful, *i.e.* converges with $\chi^2 < 225$, a B candidate is found. Each B candidate is subject to additional requirements, which include minimum p_T of the B candidate itself and the K^+, K^{*0} or ϕ daughter meson, a cut on the z position of the B vertex, and a vertex quality cut.

Accepting only certain ranges (Table 3.1) of z positions of the B vertices can be categorized as a technical quality requirement. This requirement eliminates most of the badly mis-reconstructed candidates originating from the SVX bulkhead regions. More details on how and why this cut was arrived at can be found in Appendices F and D.

The remaining cuts are categorized as "physics" cuts and are subject to cut optimization. The results of two such optimizations, for the average lifetime measurement and for $\Delta\Gamma_s/\Gamma_s$ measurement, are included in Table 3.2. The optimization procedure is described in Appendix D.

3.1.5 Other selection caveats

We have looked at the *reflections*, *i.e.* B decays incorrectly reconstructed as our signal decay, often as a result of the assignment of a wrong mass to a track — another exhibit of how a lack of PID can complicate an analysis. In the first stage of studying the reflections we reconstruct the decay of interest in the generic $b \to J/\psi X$ *realistic Monte-Carlo* sample (see Appendix C for definition and explanation), and look at the source of everything other than the true signal that ends up in the mass window using Monte-Carlo truth information. Figure 3-3 gives a qualitative picture of what is happening for the three decays we consider. In the second stage we generate and study a few decays that appear more worrisome than others. The details of these studies are described in Appendices E, H, and I.

Below we list the most prominent reflections as well as what we have done in order to prevent them from adversely affecting the results of the analyses:

- reconstructed as $B_u \to J/\psi K^+$:
  - ❶ $B_u \to J/\psi \pi^+$,
  - ❷ $B_d \to J/\psi K^{*0}$, $K^{*0} \to K^+\pi^-$ and non-resonant $B_d \to J/\psi K^+\pi^-$ with π^- missed, as well as other partially reconstructed decays.
- reconstructed as $B_d \to J/\psi K^{*0}$, $K^{*0} \to K^+\pi^-$:
  - ❸ $B_s \to J/\psi\phi$, $\phi \to K^+K^-$ with K^- assigned a π mass,
  - ❹ $B_d \to J/\psi K^+\pi^-$ non-resonant decay (with potential $K \leftrightarrow \pi$ mass mis-assignment),

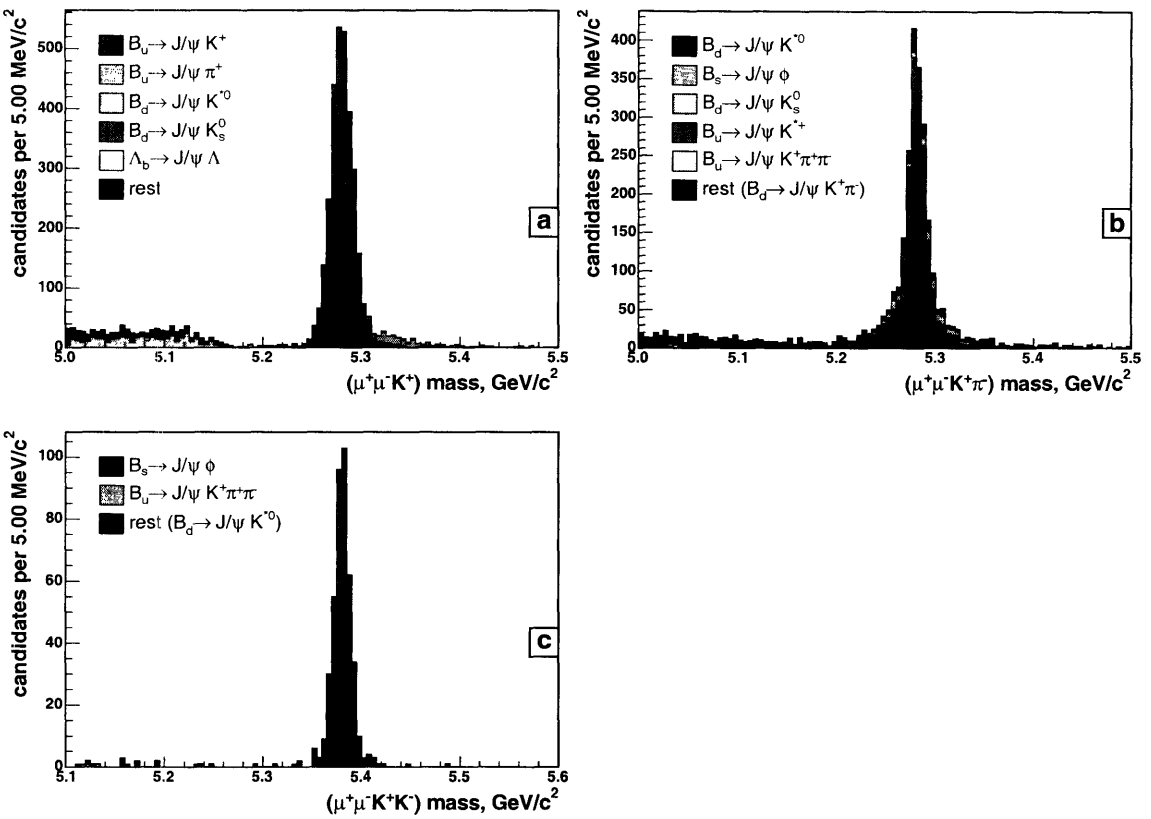


Figure 3-3: Mass spectrum when reconstructing the $B_u \to J/\psi K^+$ (a), $B_d \to J/\psi K^{*0}$ (b) and $B_s \to J/\psi\phi$ (c) decay in the $b \to J/\psi X$ Monte-Carlo.

❺ self-reflection $B_d \to J/\psi K^{*0}$, $K^{*0} \to K^+\pi^-$ with K and π mass assignments swapped.

- reconstructed as $B_s \to J/\psi\phi$, $\phi \to K^+K^-$:

❻ $B_d \to J/\psi K^{*0}$, $K^{*0} \to K^+\pi^-$ and $B_d \to J/\psi K^+\pi^-$ with the π^- assigned a K mass.

The above reflections are problematic in at least one of the following senses:

- They may distort the otherwise simple shape of combinatorial background in the mass distribution.
- They increase the amount of long-lived background (with lifetimes close to those we measure!) thus working against the decoupling of signal and the background in the proper decay length sub-space (see Section 3.2.3 for more details).
- Some of them are distorting the angular distribution of the true signal and/or making the background angular distribution look more like that of the signal.

We handle these reflections in several ways: parametrize, suppress, or eliminate them. More specifically:

- ❶ is parameterized using $B_u \to J/\psi\pi^+$ Monte-Carlo sample reconstructed as $B_u \to J/\psi K^+$. The details of this parameterization can be found in Appendix H.

- A study of ❷ shows that these reflections have masses below $5.17\,\mathrm{GeV/c^2}$, as illustrated by Figure 3-3 (a). Therefore they are eliminated by setting the mass window for B_u candidates from $5.17\,\mathrm{GeV/c^2}$ to $5.39\,\mathrm{GeV/c^2}$. The upper boundary of the mass window is set such that the window is more or less symmetric around the signal peak.

- ❸ is significantly suppressed by changing the mass window for $K\pi$ combinations from historical $M^{K^{*0}}_{PDG} \pm 80.0\,\mathrm{MeV/c^2}$ to $M^{K^{*0}}_{PDG} \pm 50.0\,\mathrm{MeV/c^2}$. This change is discussed in detail in Appendix E. The remainder of this reflection is treated as a systematic effect (Section 3.4).

- ❹ is found to be relatively small, 4% at maximum, but compatible with zero within the uncertainty. The $K\pi$ combinations that fall into K^{*0} mass window and have the correct mass assignment do not distort mass or the lifetime distribution. Those with swapped masses ($\sim$40% of the above 4%, about 1.5% of the total signal as determined from a Monte-Carlo study) have a four-track invariant mass distribution compatible with that of background and a lifetime close to that of the B_d signal. This is a very small component however, and the long-lived background component in the fit model (Section 3.2.3) should take care of it. The remaining effect of this component on the results of the measurement is quantified in Section 3.4.

- ❺ is parameterized using Monte-Carlo as discussed in Appendix I.

- ❻ is reduced by requiring a $M^{\phi}_{PDG} \pm 6.5\,\mathrm{MeV/c^2}$ mass window for ϕ candidates. The justification of this requirement and some of the implications are discussed in the Appendix E.

3.1.6 Nominal cuts

In conclusion of this section Table 3.1 gives a summary of technical cuts applied in the analysis. Table 3.2 gives a summary of physics cuts. Note, that the cuts used in the average lifetime measurement and the time-dependent angular analysis in the B_s sample are different. The difference is due to the fact that the measurement of different quantity is optimized in these two cases.

3.2 Sample composition and fitting technique

We use the unbinned maximum likelihood (MLL) fitting method [98] to extract parameters of interest from the B meson samples at hand. In this method the likelihood function $\mathcal{L}$ is defined as the product of probability density functions (PDFs) of all candidates in a given sample. The PDFs and thereby the likelihood function depend

Description	Cut value
Track quality	$p_T > 0.4\,\mathrm{GeV/c}$ (in skimming only)
Track quality	$\geqslant 3$ SVX layers with an $r-\phi$ hit
Track quality	$\geqslant 2$ axial SL with $\geqslant 5$ hits in each
Track quality	$\geqslant 2$ stereo SL with $\geqslant 5$ hits in each
Muon quality	$\geqslant 3$ axial SL with $\geqslant 5$ hits in each
Muon quality (CMU only)	$\chi^2_{CMU}(\mu^{\pm}) < 9.0$
Muon quality	$p_T(\mu^{\pm}) > 1.5\,\mathrm{GeV/c}$
J/ψ mass window	$\lvert m(\mu^+\mu^-) - M^{J/\psi}_{PDG}\rvert < 80\,\mathrm{MeV/c^2}$
z of the B vertex, cm	$\in$(-50.0; -18.0)$\cup$(-14.0; 14.5)$\cup$(18.5; 50.0)

Table 3.1: Summary of the technical quality cuts. These cuts are common to the three decay modes considered and are the same in the average lifetime and time-dependent angular analyses.

Decay	p_T^B, GeV/c	$p_T^{K^{*0}/\phi}$, GeV/c	$Prob(\chi^2)$	K^{*0}/ϕ mass window, MeV/c^2	B mass window, MeV/c^2
$B_u \to J/\psi K^+$	5.5	1.6	10^{-3}	—	5170−5390
$B_d \to J/\psi K^{*0}$	6.0	2.6	10^{-4}	$M^{K^{*0}}_{PDG} \pm 50.0$	5170−5390
$B_s \to J/\psi\phi$	5.0 / 6.0	1.5 / 2.0	10^{-5}	$M^{\phi}_{PDG} \pm 6.5$	5220−5520

Table 3.2: Summary of the B meson candidate selection for the average lifetime measurement and time-dependent angular analysis. In case of p_T cuts applied in the B_s selection, the first number is for the average lifetime measurement and the second one is for time-dependent angular analysis.

on the parameters of the model being fit. Once constructed, $\mathcal{L}$ is maximized by varying the parameters of the model and the most likely (given the data) values of the parameters are determined. Technically, the maximization procedure is performed using the `MINUIT` package [100].

Each of the B samples we are considering consists of signal and background plus, potentially, reflections which we do not explicitly include in either of the first two categories. For the signal, theory (Section 1.4) tells us which PDFs to use. For the reason that the background is mostly combinatorial, there is no particular theory for it, which means that we need to come up with an empirical model. The reflections are most straightforwardly handled by studying and parameterizing them using Monte-Carlo methods. In constructing the PDFs one should also keep in mind that the observed data have been subjected to detector and/or selection sculpting and resolution effects.

To construct the necessary PDFs, we look at the following distributions (subspaces):

- Mass, m. In mass the signal is usually a peak sitting on top of a slowly varying background shape.

- Proper decay length (PDL), ct. In PDL the signal appears as an exponential-like tail with the characteristic parameter close to $400 - 500\,\mu$m, typical for B mesons. Background is dominated by a peak at $ct = 0$ due to a large number of combinations involving prompt J/ψ, but also has some longer lived components.

- Transversity, $\vec{\omega}$ (for $B_{d,s}$ only). The transversity is defined in Section 1.4.1 and is in fact a three-dimensional sub-space. As such it is difficult to visualize or present graphically, let alone trying to describe a qualitative difference between signal and background.

The parameters we are interested in are those from the $ct \cup \vec{\omega}$ sub-space. However, we fit in the mass sub-space as well, for it is crucial for separating signal from background. The fit results for the masses, however, should *not* be taken as serious measurements, because the selection is not optimized for the mass measurement and the relevant systematic effects are not evaluated.

The bulk of the current section is devoted to detailed understanding of the components in the three B samples considered and devising PDFs for all of them. In accordance with the strategy declared in Section 1.5.4, we first develop the fit model and the corresponding PDFs for the average lifetime measurement in the three B meson samples. Next we consider a time-*independent* angular analysis of the neutral B meson decays, so that the sculpting of the angular distributions can be incorporated in the least complicated fashion. The transition to time-*dependent* angular analysis is most straightforward for B_d — because time and angular dependence fully factorize (Equation 1.85); it is just a matter of multiplication of the angular PDF by the PDF for time dependence. For B_s it is more complicated, because different angular terms in the signal have different lifetimes as Equation 1.84 implies. We complete this section by quoting the total likelihood functions constructed for all fits to the data performed in the context of this dissertation.

3.2.1 Fit variables

Once we have determined that a candidate belongs to a sample, there are only seven variables we are interested in. It is in terms of these seven the PDFs are constructed. For the candidate j these variables are:

- m_j, σ_{m_j} – mass of the candidate and its uncertainty. These are calculated as a result of the `VertexFit` intersecting B daughter tracks in a common space-point by adjusting track parameters within their uncertainties.

- ct_j, σ_{ct_j} – PDL of the candidate and its uncertainty. To measure these, three ingredients are required:

  1. Transverse momentum of the B meson candidate, $\vec{p}_T^{\,B}$, which is calculated from the momenta of the tracks comprising the candidate.
  2. (x, y) coordinates of the secondary vertex (SV), as well as the corresponding error matrix. These come from the same `VertexFit` that determines m_j, σ_{m_j}.

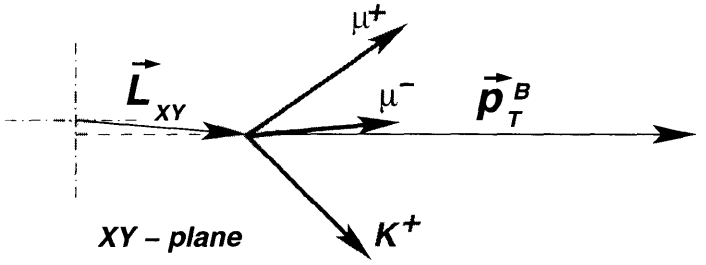


Figure 3-4: Illustration of the PDL calculation for a B_u meson candidate.

3. (x, y) coordinates of the primary vertex (PV), where the B meson was produced and the corresponding error matrix. The PV is determined as a point on a beamline (precisely measured using a large number of events from the current run) with the z position given by the z coordinate of the J/ψ candidate vertex. The PV error matrix is extracted from the measured beam profile.

With these three ingredients available, the proper decay length of a B meson candidate is calculated using

$$ct = \vec{L}_{xy} \cdot \vec{p}_T^{\,B} \frac{M_{PDG}^B}{|\vec{p}_T^{\,B}|^2}, \tag{3.1}$$

where $\vec{L}_{xy}$ is a vector in the (x, y) plane from the PV to the B decay vertex, as shown in Fig 3-4.

- $\vec{\omega}_j = (\cos\theta_j,\ \phi_j,\ \cos\psi_j)$ – transversity variables characterizing the decay of this particular candidate. Unlike the mass and the PDL, in which the distributions are not biased (or the bias is unimportant) and the major detector effect amounts to a smearing of an individual measurement with some resolution, in the transversity variables the major effect comes from sculpting, discussed in Section 3.2.4 and Appendix K, while resolutions do not appear to be important. The resolutions in each of the transversity variables have been studied using the Monte-Carlo simulation. The results of the study are depicted in Figure 3-5. Based on the following observations we decide to ignore the smearing of the transversity variables:

1. In the Monte-Carlo study, resolutions in each of the three transversity variables are found to be very small, compared to a typical size of the structure in the involved angular distributions.

2. All of the fits ignoring the resolution effects on various Monte-Carlo samples return angular decay amplitudes consistent with the input values within extremely small uncertainties.

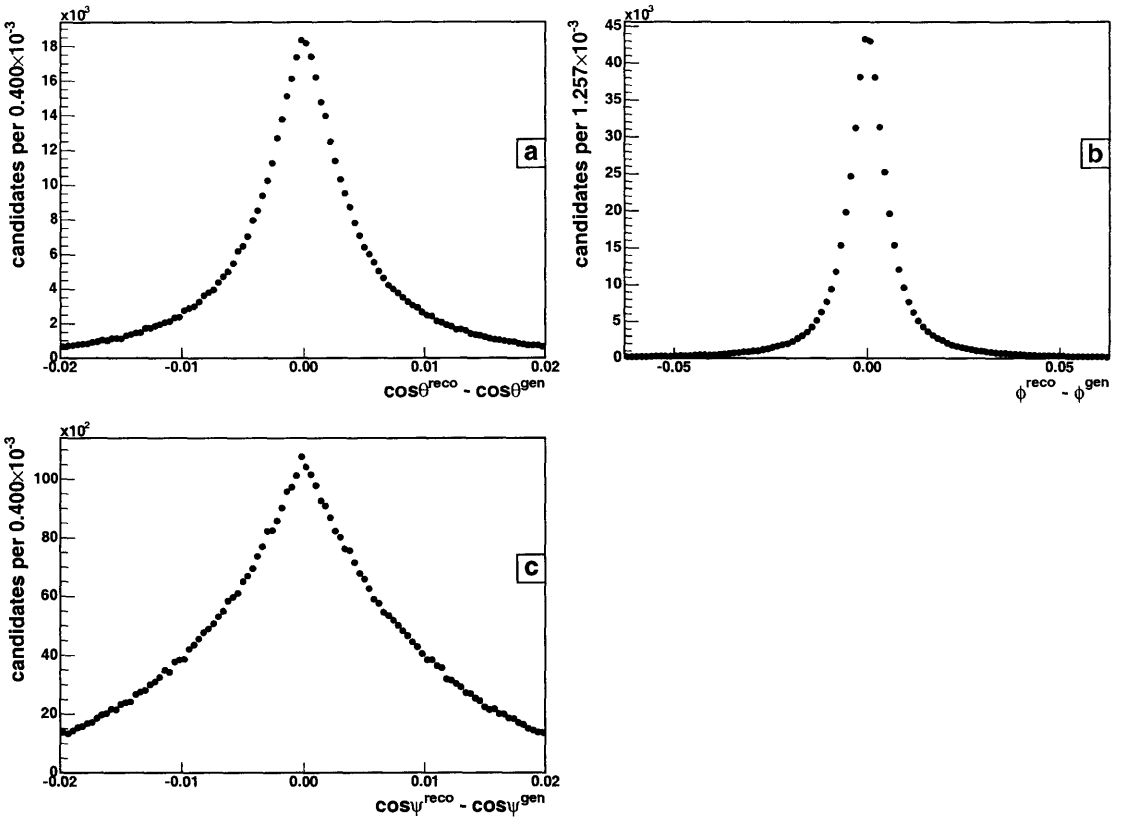


Figure 3-5: Event-by-event differences between the generated and the reconstructed $\cos\theta$, ϕ and $\cos\psi$ for a Monte-Carlo sample of $B_s \to J/\psi\phi$ decays. These plots give an idea of what the resolutions on the measured transversity variables are. The depicted range along the horizontal axis is 2% of the full range the respective variable spans.

3.2.2 PDF for the mass (m)

As a typical example of a mass distribution in the current analyses, we show the result from the B_u sample in Figure 3-6 (a).

Signal. The width of the signal peak is dominated by detector resolution, which we model by a single Gaussian. The width of the Gaussian is calculated on event-by-event basis from the uncertainty of the individual mass measurement, σ_{m_j}. These individual uncertainties can be consistently over- or under-estimated, due to the complicated nature of the tracking algorithm (see Sections 2.2.1, 3.1.1). To account for this effect we allow a global scale factor, S_m, which multiplies individual uncertainties. With the above definitions and M being the mean value for the mass of the signal, the signal PDF is

$$X_{sig}^{obs}(m_j, \sigma_{m_j}|M, S_m) = \frac{1}{\sqrt{2\pi}S_m\sigma_{m_j}} e^{-\frac{(m_j-M)^2}{2S_m^2\sigma_{m_j}^2}}. \tag{3.2}$$

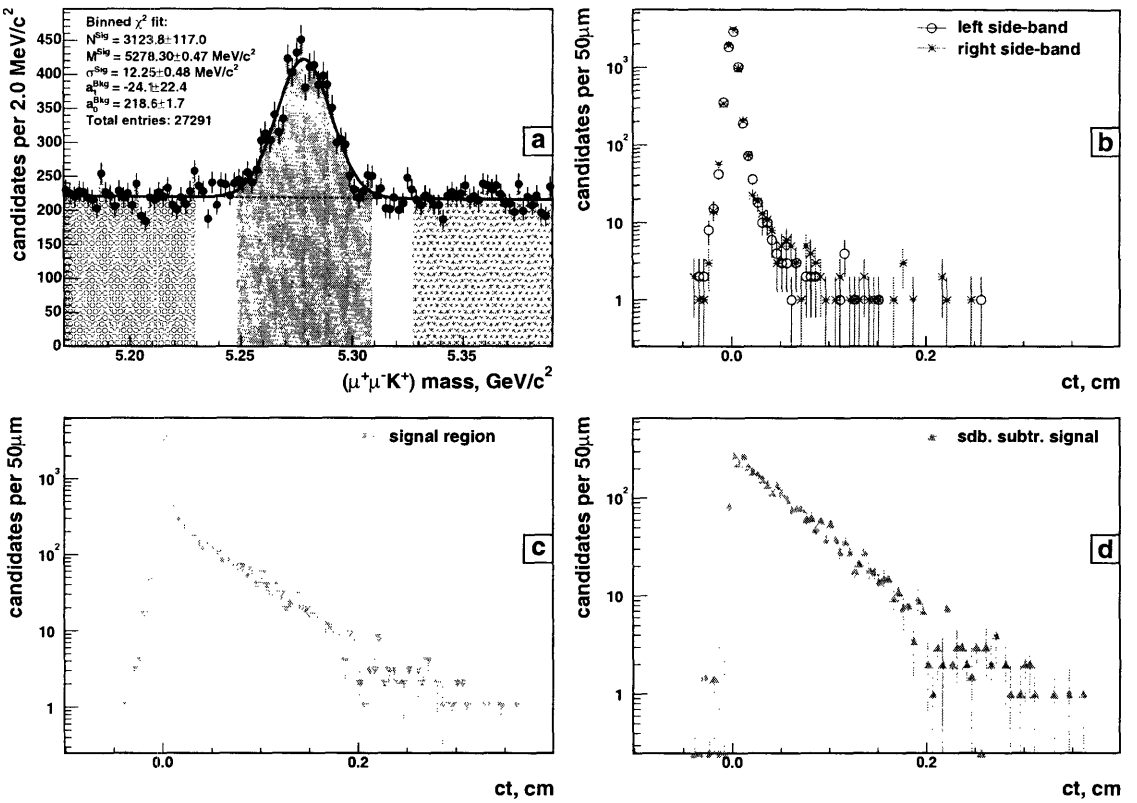


Figure 3-6: (a) mass distribution of the B_u candidates in the sample obtained with nominal cuts. The side-bands and the signal region are indicated. The binned χ^2 fit used in the side-band subtraction is superimposed. (b) PDL distribution for the left and right side-bands. (c) PDL distribution for the signal region. (d) Side-band subtracted PDL distribution for the signal.

Background. Based on the p_T spectra of generic tracks, the background shape is expected to be close to a falling exponential. However, it is not uncommon that the selection cuts introduce a fair amount of non-trivial shaping, therefore the safest way to proceed is to use an empirical parameterization. Figure 3-6 (a) suggests that a linear shape, $a\,m + b$, is a good fit. Normalizing $a\,m + b$ to unit area in the mass window from M_{min} to M_{max}, we get the following PDF for background:

$$X_{bkg}^{obs}(m_j|a) = a\,m_j + \frac{1}{M_{max} - M_{min}}\left[1 - \frac{a}{2}(M_{max}^2 - M_{min}^2)\right]. \qquad (3.3)$$

Reflections. As has been discussed in Section 3.1.5, there is a significant contribution to the $B_u \to J/\psi K^+$ sample from $B_u \to J/\psi\pi^+$ decays. A parameterization for the mass of this component is derived in Appendix H. In the $B_d \to J/\psi K^{*0}$ sample some of the signal candidates are reconstructed with K $\leftrightarrow \pi$ mass mis-assignment for K^{*0} daughter tracks. These have a shape different from that of the correctly reconstructed signal. A parameterization of this shape is devised in Appendix I.

3.2.3 PDF for the proper decay length (ct)

To devise an adequate parameterization for the PDL of the signal and the background, plots (b)–(d) in Figure 3-6 are extremely useful.

Signal. The lifetime of the signal, shown in Figure 3-6 (d), manifests itself in an exponentially decaying ct distribution smeared out by the detector resolution. The smearing can clearly be seen from the fact that some of the side-band subtracted signal has negative ct. Also approximately half of the prompt background is smeared into negative ct region, as can be seen in Figure 3-6 (b).

With detector smearing taking place, to properly extract the lifetime of the signal one needs to understand the resolution function. A single Gaussian, as discussed later in this section, appears to be an adequate description of the resolution function. The width of the Gaussian is given by the individual PDL uncertainty, σ_{ct_j}, such that the signal PDF for candidate j is given by

$$Y_{sig}^{obs}(ct_j, \sigma_{ct_j}|c\tau, S_{ct}) = E(ct_j|c\tau) \otimes G(ct_j, \sigma_{ct_j}|S_{ct}), \tag{3.4}$$

where

$$E(ct|c\tau) = \begin{cases} 0 & , \quad ct < 0 \\ \frac{1}{c\tau} e^{-\frac{ct}{c\tau}} & , \quad ct \geqslant 0, \end{cases} \tag{3.5}$$

$$G(ct, \sigma_{ct}|S_{ct}) = \frac{1}{\sqrt{2\pi} S_{ct}\sigma_{ct}} e^{-\frac{c^2t^2}{2S_{ct}^2\sigma_{ct}^2}}, \tag{3.6}$$

and τ is the lifetime we are extracting. The scale factor S_{ct} corrects consistently over- or under-estimated PDL uncertainties. An expression suitable for coding up the convolution in Equation 3.4 is given by Equation G.1.

Background. Background PDL distribution shown in Figure 3-6 (b). It appears to have a peak at zero PDL, short positively and negatively lived tails, and a longer lived tail on the positive PDL side, which suggests the following empirical parameterization:

$$\begin{aligned} &Y_{bkg}^{obs}(ct_j, \sigma_{ct_j}|f_-, f_+, f_{++}, \lambda_-, \lambda_+, \lambda_{++}, S_{ct}) \\ &= (1 - f_- - f_+ - f_{++})\, G(ct_j, \sigma_{ct_j}|S_{ct}) \\ &+ f_-\, E(-ct_j|\lambda_-) \otimes G(ct_j, \sigma_{ct_j}|S_{ct}) \\ &+ f_+\, E(ct_j|\lambda_+) \otimes G(ct_j, \sigma_{ct_j}|S_{ct}) \\ &+ f_{++}\, E(ct_j|\lambda_{++}) \otimes G(ct_j, \sigma_{ct_j}|S_{ct}). \end{aligned} \tag{3.7}$$

There are six parameters, in addition to S_{ct}, describing the shape of background ct distribution:

- λ_-, f_- – decay constant and fraction of the negatively lived tail,
- λ_+, f_+ – decay constant and fraction of the short positively lived tail,
- λ_{++}, f_{++} – decay constant and fraction of the long positively lived tail.

We convolute all of the terms in the background PDL parameterization with the resolution function, G (for Gaussian), in the same way as in Equation 3.4. This has not been done in the Run I measurement [96], which led to an unphysical discontinuity of the PDF at $ct = 0$.

Let us now, somewhat retroactively, introduce the speculations that motivate the choice of the above background PDL parameterization.

The background PDL distribution is dominated (note the logarithmic vertical scale in Figure 3-6 (b)) by a peak roughly centered at $ct = 0$. The candidates in the peak are constructed from tracks coming from the primary vertex, mostly prompt J/ψ paired with prompt track(s). The true lifetime of these candidates is mathematically described by $\delta(ct)$. The δ-function is smeared with the detector resolution, which we, in fact, assume to be the same for prompt and displaced vertices, thus the term $(1 - f_- - f_+ - f_{++})\, G(ct_j, \sigma_{ct_j}|S_{ct}) = (1 - f_- - f_+ - f_{++})\, \delta(ct_j) \otimes G(ct_j, \sigma_{ct_j}|S_{ct})$. The sheer size of the prompt contribution allows us to make reasonable assumptions about the shape of the resolution function from the B sample itself, without having to resort to any other sample. In particular, we can make a judgment about the extent to which a single Gaussian is a good description of the resolution function; *e.g.* if all σ_{ct_j} were the same, $\sigma_{ct_j} = \langle\sigma_{ct_j}\rangle\ \forall\ j$, and the resolution function were truly Gaussian then the peak in the background PDL distribution would have been Gaussian with the width $S_{ct}\langle\sigma_{ct_j}\rangle$.

Our philosophy about the short-lived $(f_\pm, \lambda_\pm)$ and the long-lived (f_{++}, λ_{++}) tails is that they should *effectively* account for the remainder of the background PDL distribution, *decoupling* it from that of the signal. The term *effectively* is used here to mean "on average", without caring too much how well a specific background contribution is described. However, there is a physics motivation for the choice of parameterization we have made.

The long-lived background may have contributions from at least some of the following sources:

- residual (not explicitly accounted for) reflections,
- true displaced J/ψ paired with a random track (one observed example of which is multiple B candidates sharing the same J/ψ),
- sequential semileptonic decays $b \to c \to s$ with resultant (μ^+, μ^-) faking a J/ψ.

These, in principle, should each have its own fraction and a corresponding lifetime, which would actually be an effective lifetime as well. However, the smallness of these contributions seen in Figure 3-6 (b) allows us to lump them together into a single fraction f_{++} with one effective lifetime λ_{++}.

The short-lived negative and positive tails handle background candidates, which:

- involve track(s) that make use of an erroneous hit,
- have tracks that in reality belong to different displaced vertices, as there are at least two in a $b\bar{b}$ event.

This latter source explains why (f_+, λ_+) component comes out of the fit larger than (f_-, λ_-), contrary to (somewhat naive) expectation of them to be more or less the same.

In any case, $(f_\pm, \lambda_\pm)$ and (f_{++}, λ_{++}) seem to make a set of components needed to adequately describe the background, without having the signal lifetime to participate. Of course, reliance on signal/background separation coming from the mass domain is absolutely essential. To see if such approach works, we look at the parameter correlation matrix obtained in each of the three fits for the average lifetime. A good sign would be if the lifetime parameter is largely decoupled from background lifetimes and fractions. Running ahead of ourselves, this is what we indeed observe. In fact, the lifetime parameter is the third least correlated parameter in the 12-parameter fit.

Finally, a systematic uncertainty is evaluated and assigned to the results due to assumptions made in describing the PDL distribution by fitting with alternative choices for the resolution function and the background PDL parameterization as discussed in Section 3.4.

Reflections. As has been pointed out in Section 3.1.5 there are significant reflections in the $B_u \to J/\psi K^+$ and $B_d \to J/\psi K^{*0}$ samples. PDFs for the proper decay length of these components are devised in Appendices H and I respectively.

3.2.4 PDF for the transversity ($\vec{\omega}$)

In this section we derive PDFs that can be used in a time-*independent* angular analysis ($B_{d,s}$ only) similar to those described in References [54–57].

With a perfect detector and no selection cuts applied, one would be able to use the trivially normalized theoretical formula (Equation 1.77 with time-dependence integrated out, which may or may not result in re-defining A_α as Appendix J discusses)

$$Z_{sig}(\vec{\omega}|\{A_\alpha\}) = \frac{1}{|A_0|^2 + |A_\parallel|^2 + |A_\perp|^2}\Big[|A_0|^2 f_1(\vec{\omega}) + |A_\parallel|^2 f_2(\vec{\omega}) + |A_\perp|^2 f_3(\vec{\omega}) \\ \pm Im(A_\parallel^* A_\perp) f_4(\vec{\omega}) + Re(A_0^* A_\parallel) f_5(\vec{\omega}) \pm Im(A_0^* A_\perp) f_6(\vec{\omega})\Big] \qquad (3.8)$$

as the PDF. However, unlike the case of ct, CDF does not have uniform sensitivity in the $\vec{\omega}$ sub-space, which means that the angular distributions we observe are *not* the true underlying ones, but rather the sculpted versions of those. To illustrate this Figure 3-7 shows the sculpting of the transversity variables in a sample of $B_s \to J/\psi\phi$ decays obtained from the Monte-Carlo simulation. Also, the cuts applied in the B meson selection procedure are specifically designed to leave the ct, but *not* the transversity, distribution of the signal unbiased, so at least some of the sculpting comes as a result of application of cuts. For example, $\cos\theta$ distribution is nearly flat (sculpting is less pronounced) for high p_T B mesons, while for low p_T ones it has a clear arch-like shape, as can be seen in Figure 3-8. The sculpting may also depend on other kinematic variables, collectively named $\vec{\kappa}$, via non-trivial detector sensitivity to them.

Through the rest of this dissertation we use the words "efficiency" and "sculpting"

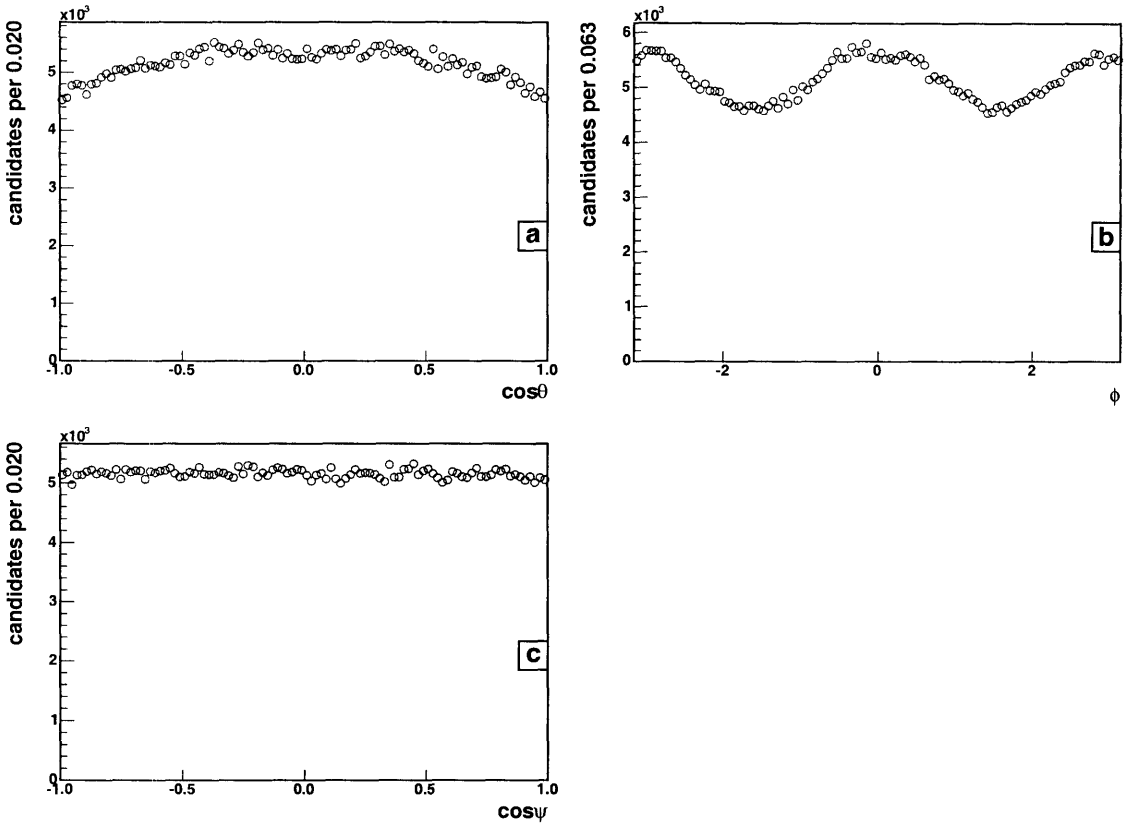


Figure 3-7: (a) $\cos\theta$, (b) ϕ and (c) $\cos\psi$ distribution of $B_s \to J/\psi\phi$ candidates reconstructed in the Monte-Carlo sample with the nominal analysis requirements. All three distributions have been generated flat.

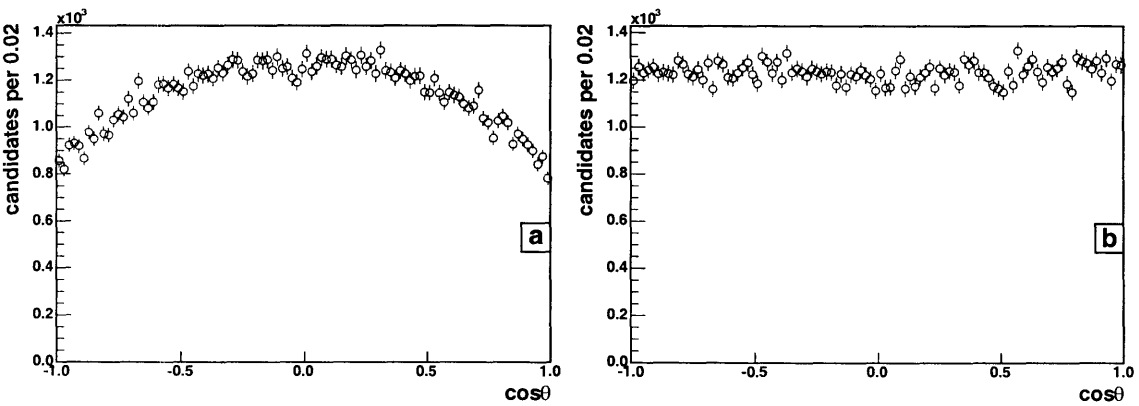


Figure 3-8: $\cos\theta$ distribution from a Monte-Carlo sample of $B_s \to J/\psi\phi$ decays: (a) $6\,\text{GeV/c} < p_T^{B_s} < 8\,\text{GeV/c}$, (b) $p_T^{B_s} > 15\,\text{GeV/c}$.

interchangeably to mean one and the same thing, which we denote $\epsilon(\vec{\omega}, \vec{\kappa})$. For a given decay mode $\epsilon(\vec{\omega}, \vec{\kappa})$ can be determined via the Monte-Carlo simulation. In this case

one can use the following definition of efficiency

$$\epsilon(\vec{\omega},\vec{\kappa}) \equiv \frac{n^{obs}(\vec{\omega},\vec{\kappa})d\vec{\omega}d\vec{\kappa}}{n^{gen}(\vec{\omega},\vec{\kappa})d\vec{\omega}d\vec{\kappa}}, \tag{3.9}$$

where $n^{gen}(\vec{\omega},\vec{\kappa})$ is the event density generated in the small phase-space volume $d\vec{\omega}d\vec{\kappa}$ centered at $(\vec{\omega},\vec{\kappa})$ and $n^{obs}(\vec{\omega},\vec{\kappa})$ is the event density observed in the same phase-space volume after detector and trigger simulation are applied and the reconstruction and the selection requirements are imposed. If one generates a sample of the decays of interest flat in $\vec{\omega}$, *e.g.* using the decay program `EvtGen` in conjunction with phase-space decay model (Appendix C), then $n^{gen}(\vec{\omega},\vec{\kappa}) = \frac{1}{8\pi}n^{gen}(\vec{\kappa})$ and Equation 3.9 simplifies to

$$\epsilon(\vec{\omega},\vec{\kappa}) = 8\pi\,\frac{n^{obs}(\vec{\omega},\vec{\kappa})}{n^{gen}(\vec{\kappa})}, \tag{3.10}$$

which makes Monte-Carlo samples generated flat in $\vec{\omega}$ useful in studying the sculpting, especially if one can afford to integrate the $\vec{\kappa}$ dependence out.

The efficiency $\epsilon(\vec{\omega},\vec{\kappa})$ is definitely a rather intricate function for which a universal analytical parameterization is unlikely to be found. Using a three-dimensional histogram to represent this function of three variables like Belle [57] may result in severe binning effects and inevitability of performing a hefty numerical integration for each candidate in each step of the log-likelihood maximization procedure. It turns out however, that transversity PDFs both for the signal and the background can be constructed in such a way that this difficulty is circumvented.

Signal. Consider a general normalized distribution given by Equation 3.8. It does not matter which of the two neutral B meson decays under study it is applied to. For convenience and in order to introduce the true angular density function, $U(\vec{\omega})$, we re-write it in the abbreviated form:

$$U(\vec{\omega}) \equiv Z_{sig}(\vec{\omega}|\{A_\alpha\}) = \sum_{i=1}^{6} A_i f_i(\vec{\omega}). \tag{3.11}$$

If Equation 3.11 is the true angular distribution Nature has built into the decay, $V(\vec{\kappa})$ is the kinematic density function, and $\epsilon(\vec{\omega},\vec{\kappa})$ is the efficiency of registering such decays, then the (normalized to 1) probability to observe a candidate in the small volume $d\vec{\omega}d\vec{\kappa}$ centered at $(\vec{\omega},\vec{\kappa})$, *i.e.* the PDF, can be written as:

$$Z^{obs}_{sig}(\vec{\omega},\vec{\kappa}|\{A_\alpha\}) = U(\vec{\omega})V(\vec{\kappa})\epsilon(\vec{\omega},\vec{\kappa})/\langle\epsilon\rangle = \sum_{i=1}^{6} A_i f_i(\vec{\omega})V(\vec{\kappa})\epsilon(\vec{\omega},\vec{\kappa})/\langle\epsilon\rangle. \tag{3.12}$$

We are actually forced to write a "two"-dimensional PDF, because $\epsilon(\vec{\omega},\vec{\kappa})$ entangles $\vec{\omega}$ and $\vec{\kappa}$ and we cannot integrate out the $\vec{\kappa}$ dependence analytically. However, this integration is all that is needed to go back to the PDF in just the $\vec{\omega}$ sub-space. In

Equation 3.12 $\langle\epsilon\rangle$ provides the normalization:

$$\langle\epsilon\rangle \equiv \iint\limits_{\omega\ \kappa} d\vec{\omega}d\vec{\kappa}\ U(\vec{\omega})V(\vec{\kappa})\epsilon(\vec{\omega},\vec{\kappa}) = \iint\limits_{\omega\ \kappa} d\vec{\omega}d\vec{\kappa}\ \Big[\sum_{i=1}^{6} A_i f_i(\vec{\omega})V(\vec{\kappa})\epsilon(\vec{\omega},\vec{\kappa})\Big]$$
$$= \sum_{i=1}^{6} A_i \underbrace{\Big[\int\limits_{\omega} d\vec{\omega}\ f_i(\vec{\omega}) \int\limits_{\kappa} d\vec{\kappa}\ V(\vec{\kappa})\epsilon(\vec{\omega},\vec{\kappa})\Big]}_{\xi_i \text{ by MC}} \equiv \sum_{i=1}^{6} A_i\xi_i\,. \tag{3.13}$$

The last step immediately above defines the partial normalization constants, ξ_i, which can be determined numerically. Indeed, calculating the "ξ_i by MC" integral is a matter of straightforward application of the original numerical Monte-Carlo integration method [99]. According to this method, if one generates a Monte-Carlo sample of decays *uniformly* in the $\vec{\omega}$ sub-space, then in the limiting case of infinite Monte-Carlo statistics, assuming Monte-Carlo reproduces $V(\vec{\kappa})$,

$$\xi_i = \frac{1}{N^{rec}} \sum_{j=1}^{N^{rec}} f_i(\vec{\omega}_j). \tag{3.14}$$

In the above equation the sum is over only the N^{rec} events that pass the trigger, the reconstruction, and the selection requirements, which collectively emulate the $\epsilon(\vec{\omega},\vec{\kappa})$ term.

Once the ξ_i are determined, the logarithm of the likelihood function for a sample of N signal candidates is:

$$\log\mathcal{L} = \log\Big[\prod_{j=1}^{N} Z_{sig}^{obs}(\vec{\omega}_j,\vec{\kappa}_j|\{A_\alpha\})\Big]$$
$$= \sum_{j=1}^{N} \log\Big[\sum_{i=1}^{6} A_i f_i(\vec{\omega}_j)V(\vec{\kappa}_j)\epsilon(\vec{\omega}_j,\vec{\kappa}_j)/\langle\epsilon\rangle\Big]$$
$$= \sum_{j=1}^{N} \log\Big[\sum_{i=1}^{6} A_i f_i(\vec{\omega}_j)\Big] - \sum_{j=1}^{N} \log\Big[\sum_{i=1}^{6} A_i\xi_i\Big] + \sum_{j=1}^{N} \log\big[V(\vec{\kappa}_j)\epsilon(\vec{\omega}_j,\vec{\kappa}_j)\big]. \tag{3.15}$$

The last term in Equation 3.15 does not depend on $\{A_\alpha\}$, the amplitudes we want to measure. It does not change when $\log\mathcal{L}$ is maximized with respect to $\{A_\alpha\}$ and therefore it can be dropped from the maximization procedure altogether. With this overall constant dropped, the function to maximize becomes

$$\log\mathcal{L} = \sum_{j=1}^{N} \log\Big[\sum_{i=1}^{6} A_i f_i(\vec{\omega}_j)\Big] - \sum_{j=1}^{N} \log\Big[\sum_{i=1}^{6} A_i\xi_i\Big], \tag{3.16}$$

in which we know everything, except the $\{A_\alpha\}$ to be determined in the maximization.

The efficiency $\epsilon(\vec{\omega}, \vec{\kappa})$ does not enter Equation 3.16. A careful inspection of this equation also shows that we only need to know ξ_i up to an overall factor, *i.e.* the factor $\frac{1}{N^{rec}}$ in Equation 3.14 is unimportant.

It is absolutely essential, however, that the Monte-Carlo simulation conducted for the purpose of calculating ξ_i does a good job in describing detector response *and* the kinematics $\vec{\kappa}$ of the decay. The quality of the Monte-Carlo simulation (it is also an issue for several other things in the analysis) is discussed in Appendix C.

Background. The information we have about the "true" background transversity distribution, if it even makes sense to speak of such thing, is fairly limited. So is our knowledge of the sculpting of it. We do, however, know how the product of the two looks like in data as demonstrated by Figure 3-9 (b)–(d).

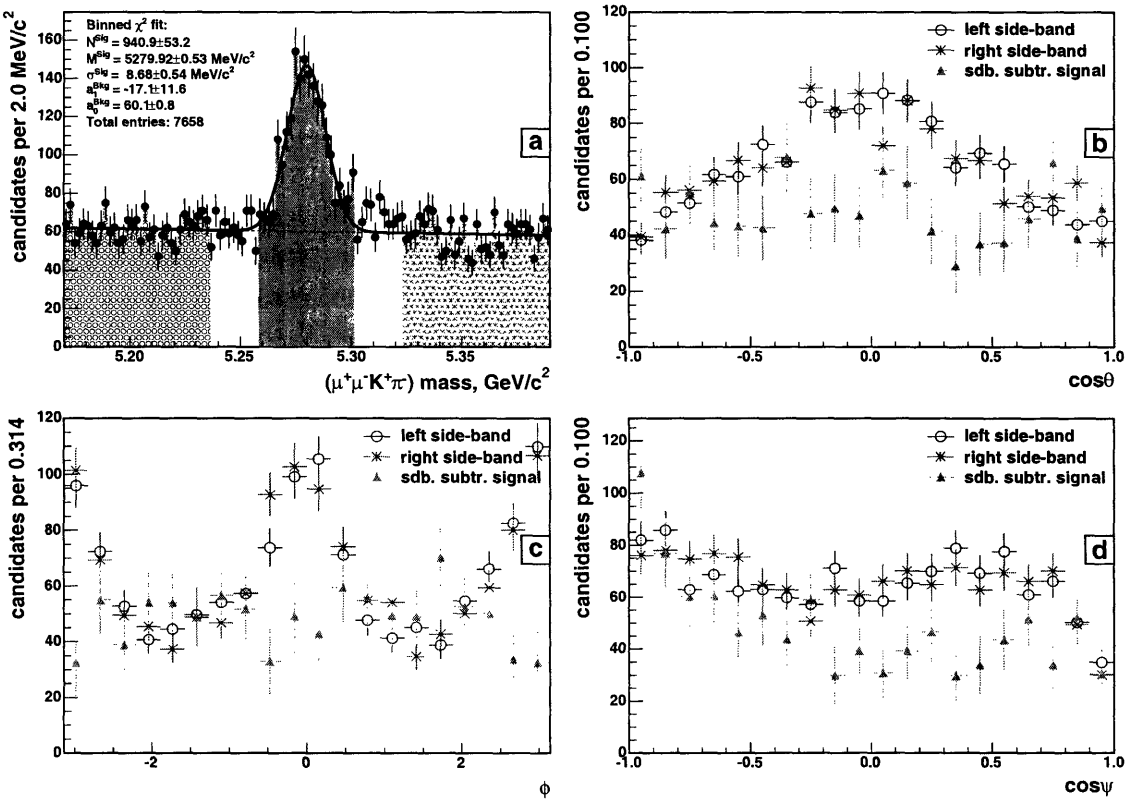


Figure 3-9: (a) mass distribution of the B_d candidates in the sample obtained with nominal cuts. The side-bands and the signal region are indicated. The binned χ^2 fit used in the side-band subtraction is superimposed. (b), (c), and (d) observed $\cos\theta$, ϕ, and $\cos\psi$ distributions for the candidates in each of the side-bands and for the side-band subtracted signal.

In such circumstances our approach to parameterizing the background transversity shape is *purely* empirical. We assume that a shape, given by Equation 3.12 with a

different set of amplitudes $\{B_\alpha\}$,

$$Z_{bkg}^{obs}(\vec{\omega},\vec{\kappa}|\{B_\alpha\}) = \frac{\sum_{i=1}^{6} B_i f_i(\vec{\omega})}{\sum_{i=1}^{6} B_i \xi_i} V(\vec{\kappa})\epsilon(\vec{\omega},\vec{\kappa}), \tag{3.17}$$

provides an adequate description. A number of comments are due at this point:

1. We need to explicitly factor out the *signal* efficiency and kinematic density from the background angular PDF, such that their product still drops out of the combined log-likelihood for signal *and* background.

2. Any well-behaved function can be projected onto a basis of orthogonal functions, whatever the basis is. This, in particular, stands true for the function that gives $\frac{Z_{bkg}^{obs}(\vec{\omega},\vec{\kappa})}{V(\vec{\kappa})\epsilon(\vec{\omega},\vec{\kappa})}$ when squared, which we choose do decompose in the partial waves basis. The decomposition may contain F or even higher partial waves, whereas Equation 3.17 allows only for S, P, and D waves. This is an artificial simplification, which requires an investigation of the systematic effects it might cause.

3. The ξ_i in Equation 3.17 are the *same* ones defined by Equation 3.13 and calculated using a *signal* Monte-Carlo sample because the six normalization integrals for Equation 3.17 are the same as the "ξ_i by MC" in Equation 3.13. Consequently, right-hand side in Equation 3.17 is normalized as long as $\sum_{\alpha=0,\|,\perp} |B_\alpha|^2 = 1$.

4. A technique similar to this was used by BaBar [56], though in a situation which much less background.

5. Left and right side-bands have very similar angular projections as can be seen in Figure 3-9 (b)–(d). This suggests that there is no strong correlation between the mass of a background candidate and its transversity angles. Still, the background may potentially consist of multiple transversity components, much the same way it breaks down into four components in PDL. Fitting a single component, *i.e.* one set of B_α, is an approximation.

6. In Run I CDF has used [55] a sum of polynomial terms of $\cos\theta$, $\cos\psi$, $\sin\phi$, and $\sin 2\phi$ to describe the background. It requires a lot of effort to ensure that such constructed PDF is positive definite everywhere in the transversity space, whereas the functional form $\sum_{i=1}^{6} B_i f_i(\vec{\omega})$ automatically guarantees that it is the case. Also it is not quite clear how signal $V(\vec{\kappa})\epsilon(\vec{\omega},\vec{\kappa})$ is factorized out in the earlier approach.

In the background transversity PDF we set $B_{4,6} = 0$, because there is no reason for the background amplitudes that would change sign when going from particle to anti-particle. There are no particles and anti-particles in the background to begin with,

there are (mostly) just random track combinations. $B_{4,6} = 0$ also means that we are not fitting for $arg(B_\perp)$.

With the thus defined background transversity PDF, the total transversity PDF for signal *and* background can be written (combining Equation 3.12 and Equation 3.17) as follows:

$$Z^{obs}_{sig+bkg}(\vec{\omega}, \vec{\kappa}|\{A_\alpha\}, \{B_\alpha\}) = \left[f_s \frac{\sum_{i=1}^{6} A_i f_i(\vec{\omega})}{\sum_{i=1}^{6} A_i \xi_i} + (1 - f_s) \frac{\sum_{i=1,2,3,5} B_i f_i(\vec{\omega})}{\sum_{i=1,2,3,5} B_i \xi_i} \right] V(\vec{\kappa})\epsilon(\vec{\omega}, \vec{\kappa}), \quad (3.18)$$

where f_s is the fraction of the signal in the sample. The product of the signal kinematic density and efficiency, $V(\vec{\kappa})\epsilon(\vec{\omega}, \vec{\kappa})$, factors out, which means that it will only enter the log-likelihood as a separate $\sum_{j=1}^{N} \log\left[V(\vec{\kappa})\epsilon(\vec{\omega}_j, \vec{\kappa}_j)\right]$ term, which, again, can be ignored in the maximization procedure.

$K \leftrightarrow \pi$ mass mis-assignment in $K^{*0} \to K^+\pi^-$ from $B_d \to J/\psi K^{*0}$. Another challenge arises in the case of $B_d \to J/\psi K^{*0}$ decay from the fact that in about 10% of the signal candidates we swap the correct mass assignments of the tracks coming from the K^{*0} decay. As a result, B_d candidates reconstructed with this pathology end up being distributed quite differently (Figure 3-10) in the *observed* $\vec{\omega}$ space from the correctly reconstructed ones, while both come from the same true underlying angular distribution with $\{A_\alpha\}$. In addition, for such mis-reconstructed B_d candidates certain selection cuts, such as $p_T^{K^{*0}}$ and $p_T^{B_d}$, are not applied on the correctly calculated quantities, which results in a different sculpting. Clearly, if special care is not taken, these *swapped* candidates will distort $\{A_\alpha\}$ extracted from the sample.

In order to achieve the desired simplification of the log-likelihood, we need to be able to factor out one common efficiency times kinematic density for the *unswapped* and *swapped* components of the signal. However, if we do not want to give up the idea of using both components to measure $\{A_\alpha\}$, we cannot settle for an empirical description of the swapped component.

The issue boils down, as one might already suspect, to being able to find an analytical expression for the product of the ratio of the efficiencies and the ratio of the kinematic densities for the swapped and unswapped components, $\frac{V^{sw}(\vec{\kappa})}{V^{un}(\vec{\kappa})} \frac{\epsilon^{sw}(\vec{\omega},\vec{\kappa})}{\epsilon^{un}(\vec{\omega},\vec{\kappa})}$. To solve this problem it helps to think in terms of a resolution function. Let $\rho' = (\vec{\omega}', \vec{\kappa}')$ be the true angular and kinematic variables of the decay, and $\rho = (\vec{\omega}, \vec{\kappa})$ – the observed ones. Let $F(\rho') = U(\vec{\omega}')V(\vec{\kappa}')$ be the density function in the ρ' space and think of a sample of events from the Monte-Carlo simulation that implements this density function. The number of candidates reconstructed at ρ is the convolution

$$N(\rho) = N^{gen} \int d\rho' F(\rho') P(\rho|\rho'), \quad (3.19)$$

where $P(\rho|\rho')$ is the probability of reconstructing a decay at ρ given that it was

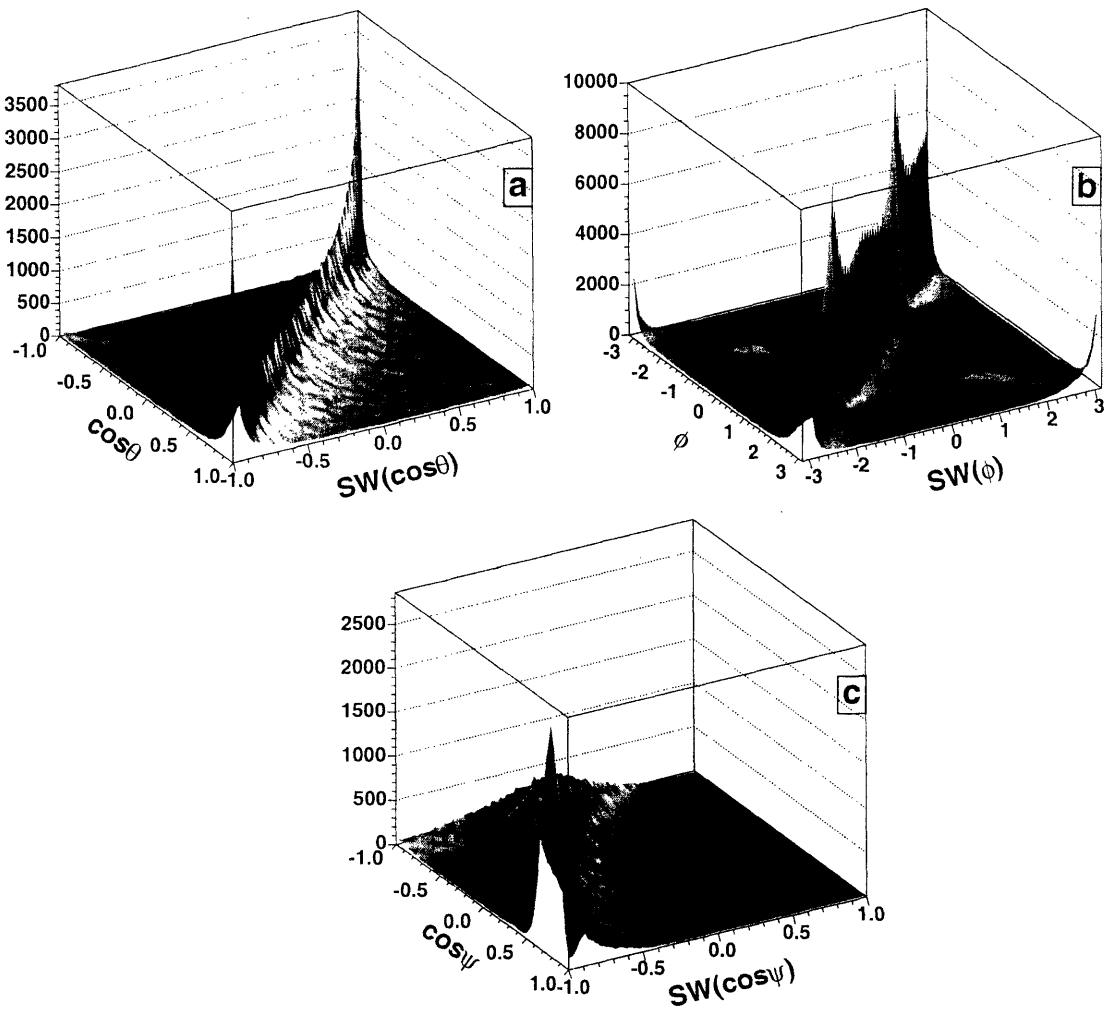


Figure 3-10: Transversity variable calculated with SWapped mass assignment in $K^{*0} \to K^+\pi^-$ reconstruction vs. the one calculated correctly: (a) $\cos\theta$, (b) ϕ, (c) $\cos\psi$. $B_d \to J/\psi K^{*0}$ Monte-Carlo sample is used to prepare these plots.

generated at ρ' and the total of N^{gen} decays were generated.

For the correctly reconstructed candidates, *i.e.* unswapped component of the signal, we, ignoring resolution effects, assume $\rho = \rho'$, *i.e.*

$$P(\rho|\rho') = \delta(\rho - \rho')\epsilon^{un}(\rho, \rho'), \tag{3.20}$$

where $\epsilon^{un}(\rho, \rho')$ is just another way of writing efficiency $\epsilon^{un}(\vec{\omega}, \vec{\kappa})$ for the unswapped candidates, which we adopt temporarily. Substituting Equation 3.20 into Equation 3.19 we get

$$N^{un}(\rho) = N^{gen} \int d\rho' F(\rho')\delta(\rho - \rho')\epsilon^{un}(\rho, \rho') = N^{gen} F(\rho)\epsilon^{un}(\rho, \rho). \tag{3.21}$$

Next, let us look at the swapped candidates. For these the probability function is

$$P(\rho|\rho') = \delta(\rho - \mathcal{SW}(\rho'))\epsilon^{sw}(\rho,\rho'), \tag{3.22}$$

that is, the observed ρ is the one we get if we swap the correct mass assignments in the K^{*0} reconstruction for a candidate with true ρ'. Again, $\epsilon^{sw}(\rho,\rho')$ is, by definition, the efficiency $\epsilon^{sw}(\vec{\omega},\vec{\kappa})$ of the swapped events. $\mathcal{SW}$ is, obviously, a well-behaved one-to-one mapping of the ρ space onto itself.

The number of reconstructed swapped candidates at ρ is

$$N^{sw}(\rho) = N^{gen}\int d\rho' \overline{F}(\rho')\delta(\rho - \mathcal{SW}(\rho'))\epsilon^{sw}(\rho,\rho'). \tag{3.23}$$

We need $\overline{F}(\rho') = \overline{U}(\vec{\omega}')V(\vec{\kappa}')$ here because for a swapped candidate the underlying angular distribution is the one for the charge conjugated particle. Unlike the case of correctly reconstructed candidates, the δ-function is now a function of $\mathcal{SW}(\rho')$, not the integration variable ρ'. So we apply the usual rule, $\delta(f(x)) = \frac{1}{|f'(x)|}\delta(x)$, which in multidimensional case takes the form:

$$\delta(\rho - \mathcal{SW}(\rho')) = \frac{1}{|\mathcal{J}(\rho')|}\delta(\mathcal{SW}^{-1}(\rho) - \rho') = \frac{1}{|\mathcal{J}(\rho')|}\delta(\mathcal{SW}(\rho) - \rho'), \tag{3.24}$$

in which $\mathcal{J}(\rho')$ is the Jacobian of the $\mathcal{SW}$ transformation. In the last step above we made use of the fact that $\mathcal{SW}^{-1} = \mathcal{SW}$. Now we can perform the integration in Equation 3.23:

$$N^{sw}(\rho) = N^{gen}\overline{F}(\mathcal{SW}(\rho))\frac{\epsilon^{sw}(\rho,\mathcal{SW}(\rho))}{|\mathcal{J}(\mathcal{SW}(\rho))|}. \tag{3.25}$$

It is now useful to re-write Equations 3.21 and 3.25 in terms of U and V:

$$N^{un}(\vec{\omega},\vec{\kappa}) = N^{gen}U(\vec{\omega})V(\vec{\kappa})\epsilon^{un}(\vec{\omega},\vec{\kappa}),, \tag{3.26}$$

$$N^{sw}(\vec{\omega},\vec{\kappa}) = N^{gen}\frac{1}{|\mathcal{J}(\mathcal{SW}(\vec{\omega},\vec{\kappa}))|}\overline{U}(\mathcal{SW}(\vec{\omega}))V(\mathcal{SW}(\vec{\kappa}))\epsilon^{sw}(\vec{\omega},\vec{\kappa}). \tag{3.27}$$

In the last equation $\mathcal{SW}(\vec{\omega})$ means transversity variables obtained by swapping the kinematics $\vec{\kappa}$.

One can generate a specific Monte-Carlo sample with known angular distribution (we take a sample with flat angular density $U(\vec{\omega}) = \overline{U}(\mathcal{SW}(\vec{\omega})) = 1$ as an example, thus the subscript $_F$) and use Equations 3.26 and 3.27 to determine the ratio

$$R(\vec{\omega},\vec{\kappa}) \equiv \frac{N_F^{sw}(\vec{\omega},\vec{\kappa})}{N_F^{un}(\vec{\omega},\vec{\kappa})} = \frac{1}{|\mathcal{J}(\mathcal{SW}(\vec{\omega},\vec{\kappa}))|}\frac{V(\mathcal{SW}(\vec{\kappa}))}{V(\vec{\kappa})}\frac{\epsilon^{sw}(\vec{\omega},\vec{\kappa})}{\epsilon^{un}(\vec{\omega},\vec{\kappa})} \tag{3.28}$$

in terms of the observed variables. One can think of $R(\vec{\omega},\vec{\kappa})$ as the ratio of the two multidimensional histograms, $N_F^{un}(\vec{\omega},\vec{\kappa})$ and $N_F^{sw}(\vec{\omega},\vec{\kappa})$, coming out of the Monte-Carlo sample used. The $\cos\theta$, ϕ, $\cos\psi$, and $p_T^{B_d}$ projections of $N_F^{un}(\vec{\omega},\vec{\kappa})$ and $N_F^{sw}(\vec{\omega},\vec{\kappa})$ are shown correspondingly in Figures 3-11 and 3-12.

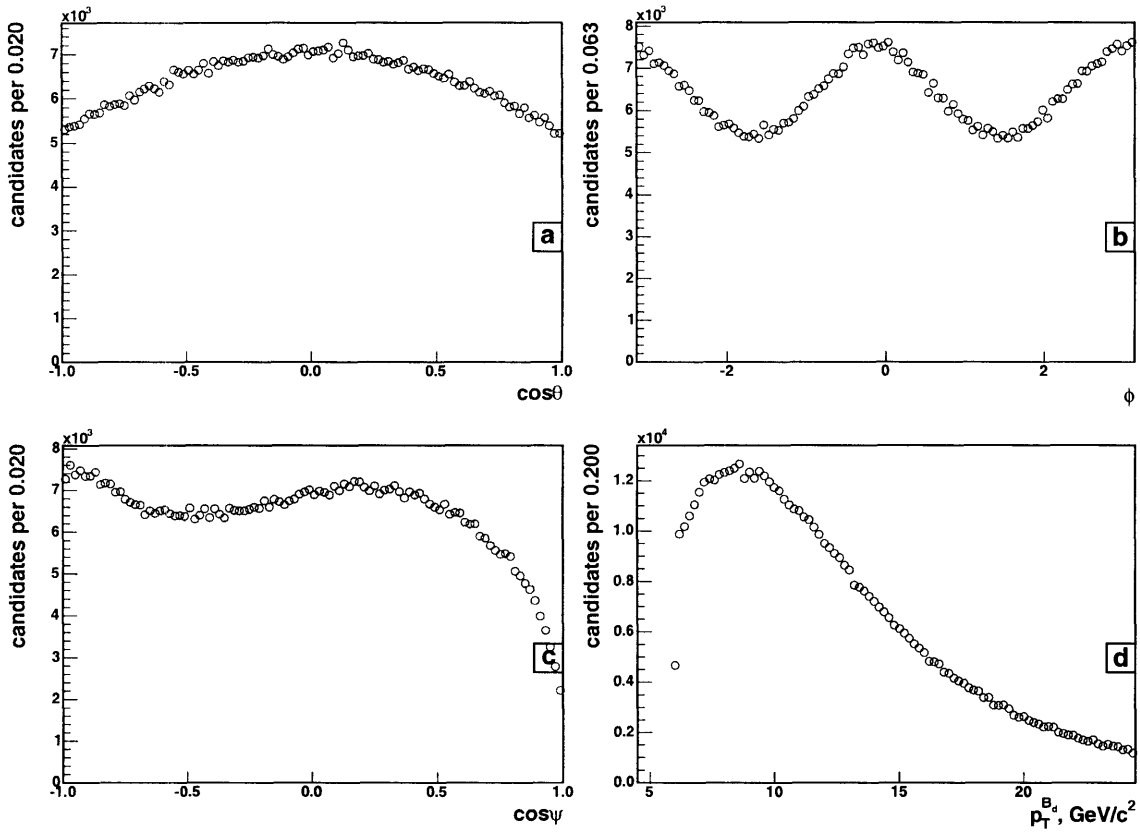


Figure 3-11: (a) $\cos\theta$, (b) ϕ, (c) $\cos\psi$, and (d) $p_T^{B_d}$ distributions of $B_d \to J/\psi K^{*0}$ candidates reconstructed in a Monte-Carlo sample with the nominal analysis requirements. Monte-Carlo truth information is used to ensure that the mass assignment in $K^{*0} \to K^+\pi^-$ reconstruction is done correctly. The three angular variable distributions have been generated flat.

Turning for a second to a practical side of determining $R(\vec{\omega},\vec{\kappa})$ described in Appendix K, we observe that it depends on $\vec{\kappa}$ only rather weakly. The *approximation*

$$\frac{1}{|\mathcal{J}(\mathcal{SW}(\vec{\omega},\vec{\kappa}))|}\frac{V(\mathcal{SW}(\vec{\kappa}))}{V(\vec{\kappa})}\frac{\epsilon^{sw}(\vec{\omega},\vec{\kappa})}{\epsilon^{un}(\vec{\omega},\vec{\kappa})} = R(\vec{\omega},\vec{\kappa}) \approx R(\vec{\omega}) \equiv \frac{N_F^{sw}(\vec{\omega})}{N_F^{un}(\vec{\omega})} \tag{3.29}$$

allows for a fairly straightforward parameterization of the left-hand side of Equation 3.29 using the Monte-Carlo simulation. Our choice to ignore $\vec{\kappa}$ dependence of $R(\vec{\omega},\vec{\kappa})$ is treated as a systematic effect in Section 3.4.5.

Let us now apply Equations 3.26 and 3.27 to the actual B_d data sample; N^{gen} now has a meaning of the total number of $B_d \to J/\psi K^{*0}$ decays occurred in the detector. Putting unswapped and swapped terms together, one has the following chance of

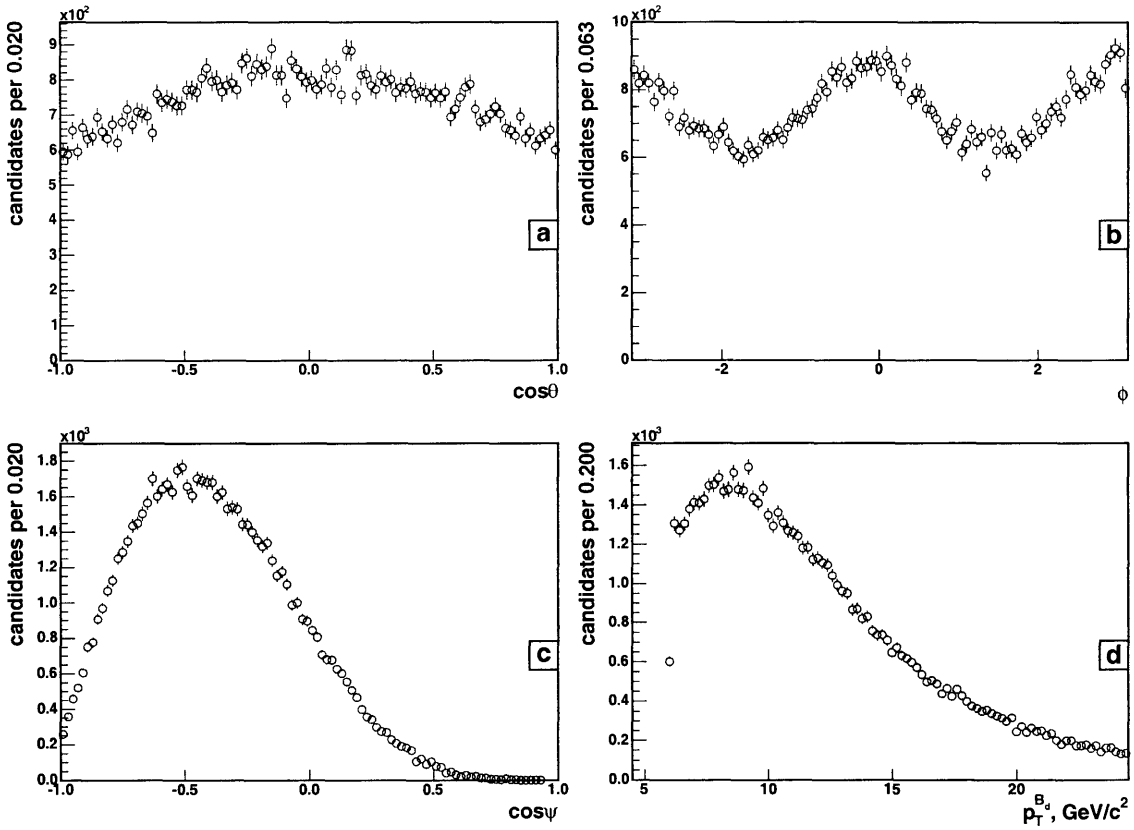


Figure 3-12: (a) $\cos\theta$, (b) ϕ, (c) $\cos\psi$, and (d) $p_T^{B_d}$ distributions of $B_d \to J/\psi K^{*0}$ candidates reconstructed in a Monte-Carlo sample with the nominal analysis requirements. Monte-Carlo truth information is used to make sure that only the candidates with mass mis-assignment in $K^{*0} \to K^+\pi^-$ reconstruction contribute to the plots. The three angular variable distributions have been generated flat.

reconstructing a signal candidates at $(\vec{\omega}, \vec{\kappa})$:

$$\begin{aligned}
\frac{N^{sig}(\vec{\omega},\vec{\kappa})}{N^{gen}} &= \frac{N^{un}(\vec{\omega},\vec{\kappa}) + N^{sw}(\vec{\omega},\vec{\kappa})}{N^{gen}} \\
&= \left[U(\vec{\omega})V(\vec{\kappa})\epsilon^{un}(\vec{\omega},\vec{\kappa}) + \frac{\overline{U}(\mathcal{SW}(\vec{\omega}))V(\mathcal{SW}(\vec{\kappa}))}{|\mathcal{J}(\mathcal{SW}(\vec{\omega},\vec{\kappa}))|}\epsilon^{sw}(\vec{\omega},\vec{\kappa}) \right] \\
&= \left[U(\vec{\omega}) + \frac{\overline{U}(\mathcal{SW}(\vec{\omega}))}{|\mathcal{J}(\mathcal{SW}(\vec{\omega},\vec{\kappa}))|}\frac{V(\mathcal{SW}(\vec{\kappa}))}{V(\vec{\kappa})}\frac{\epsilon^{sw}(\vec{\omega},\vec{\kappa})}{\epsilon^{un}(\vec{\omega},\vec{\kappa})} \right] V(\vec{\kappa})\epsilon^{un}(\vec{\omega},\vec{\kappa}) \\
&= \Big[U(\vec{\omega}) + \overline{U}(\mathcal{SW}(\vec{\omega}))R(\vec{\omega}) \Big] V(\vec{\kappa})\epsilon^{un}(\vec{\omega},\vec{\kappa}),
\end{aligned} \tag{3.30}$$

where in the last step we use the approximation of Equation 3.29.

Equation 3.30 is expressed entirely in terms of the observed variables and describes within the approximations made the observed angular distribution of the entire B_d signal. This means that after appropriate normalization it can be used as the PDF.

To obtain the normalization we need to calculate the integral:

$$\iint\limits_{\omega\ \kappa} d\vec{\omega}d\vec{\kappa}\ A_i\Big[f_i(\vec{\omega}) + \overline{f}_i(\mathcal{SW}(\vec{\omega}))R(\vec{\omega})\Big]V(\vec{\kappa})\epsilon^{un}(\vec{\omega},\vec{\kappa}) \equiv \sum_{i=1}^{6} A_i\xi_i^{un} + A_i\xi_i^{sw}, \quad (3.31)$$

which can be computed *á la* Equation 3.14 by summing over the *unswapped* candidates reconstructed from a Monte-Carlo sample of $B_d \to J/\psi K^{*0}$ decays generated flat in $\vec{\omega}$:

$$\xi_i^{un} = \frac{1}{N^{rec,un}} \sum_{j=1}^{N^{rec,un}} f_i(\vec{\omega}_j), \quad (3.32)$$

$$\xi_i^{sw} = \frac{1}{N^{rec,un}} \sum_{j=1}^{N^{rec,un}} f_i(\mathcal{SW}(\vec{\omega}_j))R(\vec{\omega}_j). \quad (3.33)$$

Again, the summation in Equations 3.32 and 3.33 is only over the *unswapped* candidates, because it is $V(\vec{\kappa})\epsilon^{un}(\vec{\omega},\vec{\kappa})$ that we factor out in Equation 3.31.

With ξ_i^{un} and ξ_i^{sw} determined, the complete transversity PDF (including the background) for the B_d sample is:

$$Z^{obs,B_d}_{un+sw+bkg}(\vec{\omega},\vec{\kappa}|\{A_\alpha\},\{B_\alpha\}) = \Bigg[f_s \frac{\sum_{i=1}^{6} A_i\big[f_i(\vec{\omega}) + \overline{f}_i(\mathcal{SW}(\vec{\omega}))R(\vec{\omega})\big]}{\sum_{i=1}^{6} A_i(\xi_i^{un} + \xi_i^{sw})} + (1-f_s)\frac{\sum_{i=1,2,3,5} B_i f_i(\vec{\omega})}{\sum_{i=1,2,3,5} B_i\xi_i^{un}}\Bigg] V(\vec{\kappa})\epsilon^{un}(\vec{\omega},\vec{\kappa}), \quad (3.34)$$

and $V(\vec{\kappa})\epsilon^{un}(\vec{\omega},\vec{\kappa})$ still drops out of the log-likelihood function to be maximized.

3.2.5 Fit functions used in the analysis

Now we have all the bits and pieces to construct the multivariate PDFs that we use in the analyses of this dissertation. For the time-dependent angular analysis however, it makes more sense to write out not the PDFs, but the $\log\mathcal{L}$ functions, so that the confusion from not explicitly known efficiencies is avoided.

Average lifetime measurement. The combined PDF is direct product of the PDF for mass (Section 3.2.2) and the PDF for proper decay length (Section 3.2.3) because there is no correlation between the two quantities for either signal or background:

$$\begin{aligned}&\mathrm{PDF}(m_j,\sigma_{m_j},ct_j,\sigma_{ct_j}|M,S_m,a,f_s,c\tau,S_{ct},f_-,f_+,f_{++},\lambda_-,\lambda_+,\lambda_{++})\\ &\quad = f_s\, X^{obs}_{sig}(m_j,\sigma_{m_j}|\ldots)\, Y^{obs}_{sig}(ct_j,\sigma_{ct_j}|\ldots)\\ &\quad + (1-f_s)\, X^{obs}_{bkg}(m_j|\ldots)\, Y^{obs}_{bkg}(ct_j,\sigma_{ct_j}|\ldots)\,. \end{aligned} \quad (3.35)$$

In the above equation f_s is the fraction of the signal in the sample. We also took the liberty of replacing the lengthy parameter lists in the right-hand side of Equation 3.35 by triple dots for the sake of conciseness and clarity.

Time-dependent angular analysis in $B_d \to J/\psi K^{*0}$. According to Equation 1.85 all signal angular terms have the same time dependency, given by the average B_d lifetime, therefore the combined signal PDF is constructed simply as a product of the mass term, PDL term and the angular term. Again, from the first principles there is no correlation among the three of them. The background description is empirical, but plotting one variable vs. another suggests no correlation either. Assuming a total of N events in the sample, the complete log-likelihood becomes:

$$\begin{aligned}\log\mathcal{L} = \sum_{j=1}^{N} \log\Bigg[& f_s \frac{\sum_{i=1}^{6} A_i f_i(\vec{\omega}_j)}{\sum_{i=1}^{6} A_i(\xi_i^{B_d,un} + \xi_i^{B_d,sw})} \\ & \times X^{obs}_{sig,un}(m_j, \sigma_{m_j}|\ldots)\, Y^{obs}_{sig}(ct_j, \sigma_{ct_j}|\ldots) \\ & + f_s \frac{\sum_{i=1}^{6} A_i \overline{f}_i(\mathcal{SW}(\vec{\omega}_j))R(\vec{\omega}_j)}{\sum_{i=1}^{6} A_i(\xi_i^{B_d,un} + \xi_i^{B_d,sw})} X^{obs}_{sig,sw}(m_j|\ldots)\, Y^{obs}_{sig}(ct_j, \sigma_{ct_j}|\ldots) \\ & + (1-f_s) \frac{\sum_{i=1,2,3,5} B_i f_i(\vec{\omega}_j)}{\sum_{i=1,2,3,5} B_i \xi_i^{B_d,un}} X^{obs}_{bkg}(m_j|\ldots)\, Y^{obs}_{bkg}(ct_j, \sigma_{ct_j}|\ldots)\Bigg]. \end{aligned} \tag{3.36}$$

$X^{obs}_{sig,sw}(m_j|\ldots)$ is given by Equation I.1, all other PDFs are defined in Section 3.2.2 and Section 3.2.3.

Time-dependent angular analysis in $B_s \to J/\psi\phi$. For this decay mode time dependence and angular distributions do not factorize (Equation 1.84), therefore we first need to spell out a PDF for time-dependent angular distribution.

The difference between $B_s \to J/\psi\phi$ and $B_d \to J/\psi K^{*0}$ decay channels is that some of the terms in the angular distribution of the former have the lifetime different from the one others have, thus the time-dependent angular distribution of the signal in $B_s \to J/\psi\phi$ is given by appropriate combination of Equation 3.12 and Equation 3.4:

$$\begin{aligned} & \mathrm{PDF}(\vec{\omega}_j, ct_j, \sigma_{ct_j}, \vec{\kappa}_j|\{A_\alpha\}, c\tau_L, c\tau_H, S_{ct}) \\ & = \frac{1}{\sum_{i=1,2,5} c\tau_L A_i \xi_i^{B_s} + c\tau_H A_3 \xi_3^{B_s}} \\ & \times \Big[\sum_{i=1,2,5} c\tau_L A_i f_i(\vec{\omega}_j)\; Y^{obs}_{sig}(ct_j, \sigma_{ct_j}|c\tau_L, S_{ct}) \\ & + c\tau_H A_3 f_3(\vec{\omega}_j)\; Y^{obs}_{sig}(ct_j, \sigma_{ct_j}|c\tau_H, S_{ct}) \Big] V(\vec{\kappa}_j)\epsilon(\vec{\omega}_j, \vec{\kappa}_j). \end{aligned} \tag{3.37}$$

The normalization and the origin of $c\tau_{L,H}$ factors are explained in Appendix J.

Knowing the above PDF the complete $\log\mathcal{L}$ over a sample of N candidates is straightforward:

$$\begin{aligned}\log\mathcal{L} = \sum_{j=1}^{N} \log\Bigg[& f_s \frac{X^{obs}_{sig}(m_j, \sigma_{m_j}|\ldots)}{\sum\limits_{i=1,2,5} c\tau_L A_i \xi_i^{B_s} + c\tau_H A_3 \xi_3^{B_s}} \\ & \times \Big[\sum_{i=1,2,5} c\tau_L A_i f_i(\vec{\omega}_j)\, Y^{obs}_{sig}(ct_j, \sigma_{ct_j}|c\tau_L, S_{ct}) \\ & + c\tau_H A_3 f_3(\vec{\omega}_j)\, Y^{obs}_{sig}(ct_j, \sigma_{ct_j}|c\tau_H, S_{ct}) \Big] \\ & + (1-f_s) \frac{\sum\limits_{i=1,2,3,5} B_i f_i(\vec{\omega}_j)}{\sum\limits_{i=1,2,3,5} B_i \xi_i^{B_s}} X^{obs}_{bkg}(m_j|\ldots)\, Y^{obs}_{bkg}(ct_j, \sigma_{ct_j}|\ldots) \Bigg] . \end{aligned} \tag{3.38}$$

3.3 Fit details and results

This section gives the remaining specifics and documents the results of the nominal fits performed on the B samples considered. A number of interesting variations of these fits, such as fitting with a different choice of parameters, applying constraints derived from the previous measurements, *etc.*, are also described in this section and are an essential part of the whole analysis. The results obtained here are not yet complete because they need to be supplemented with the systematic uncertainties discussed in Section 3.4.

3.3.1 Fitting for the average lifetime

$B_u \to J/\psi(\mu^+\mu^-)K^+$. To fit for the lifetime in this sample we use the PDF given by Equation 3.35, updated to include the $B_u \to J/\psi\pi^+$ contribution per discussion in Appendix H.

A total of 27291 candidates pass the selection requirements. The results of the fit performed on this sample are collected in Table 3.4. Fit projections onto the mass and the PDL distributions along with the results of the fit quality test, explained in Appendix L, are shown in Figure 3-13.

The correlation matrix returned by the fit is given in Table 3.3. It shows that while there are fairly strong correlations among the background PDL parameters, the signal parameters are quite uncorrelated.

$B_d \to J/\psi(\mu^+\mu^-)K^{*0}(K^+\pi^-)$. In the fit for the average lifetime in this decay mode we use the PDF given by Equation 3.35, which we modify so that it accounts for the signal component with K $\leftrightarrow$ π mass mis-assignment in the K^{*0} reconstruction. The details of this modification are elaborated in Appendix I.

With the nominal cuts, the sample that goes into the fit has 7658 candidates. The results of the fit are compiled in Table 3.4. Fit projections onto the mass and the PDL distributions along with the results of the fit quality test are shown in Figure 3-14.

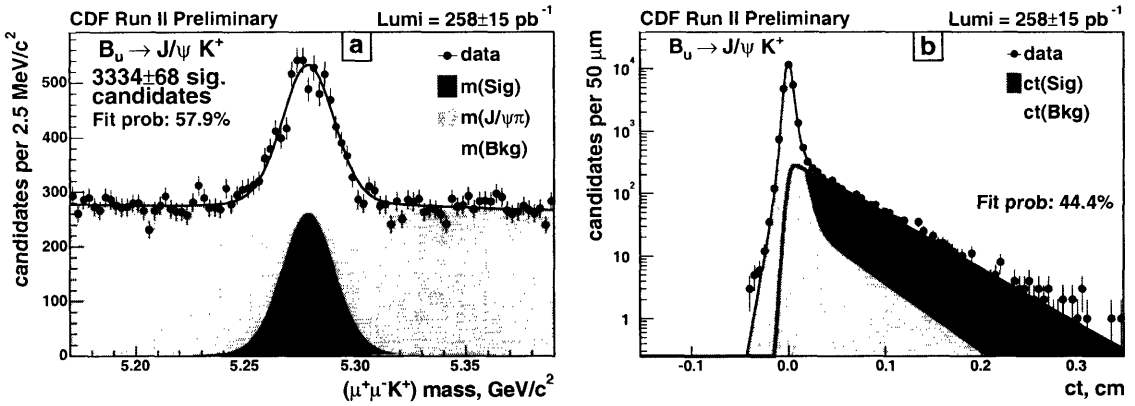


Figure 3-13: Projection of the results of the unbinned maximum likelihood fit onto (a) mass and (b) proper decay length distributions in the B_u sample.

Par.	GLOB.	M	f_s	A	S_m	$f_{J/\psi\pi}$	$c\tau$	f_-	f_+	f_{++}	λ_-	λ_+	λ_{++}	S_{ct}
M	0.0888	1	-0.039	-0.021	-0.031	-0.069	0.006	-0.005	-0.002	0.028	0.003	0.007	0.018	0.005
f_s	0.3959	-0.039	1	-0.019	0.184	0.166	-0.210	-0.007	-0.076	-0.148	0.009	-0.012	-0.023	0.015
A	0.0637	-0.021	-0.019	1	-0.001	-0.055	0.001	-0.000	0.001	0.014	-0.000	0.003	0.012	-0.000
S_m	0.3227	-0.031	0.184	-0.001	1	-0.014	-0.042	0.008	-0.003	-0.145	-0.004	-0.026	-0.048	-0.006
$f_{J/\psi\pi}$	0.3460	-0.069	0.166	-0.055	-0.014	1	-0.002	0.006	0.003	-0.125	-0.003	-0.037	-0.108	-0.003
$c\tau$	0.3190	0.006	-0.210	0.001	-0.042	-0.002	1	0.018	0.075	0.089	-0.017	-0.045	-0.188	-0.023
f_-	0.9104	-0.005	-0.007	-0.000	0.008	0.006	0.018	1	0.571	0.142	-0.852	-0.381	-0.111	-0.720
f_+	0.8485	-0.002	-0.076	0.001	-0.003	0.003	0.075	0.571	1	0.169	-0.450	-0.673	-0.107	-0.624
f_{++}	0.8152	0.028	-0.148	0.014	-0.145	-0.125	0.089	0.142	0.169	1	-0.118	-0.612	-0.652	-0.132
λ_-	0.8600	0.003	0.009	-0.000	-0.004	-0.003	-0.017	-0.852	-0.450	-0.118	1	0.305	0.090	0.531
λ_+	0.8708	0.007	-0.012	0.003	-0.026	-0.037	-0.045	-0.381	-0.673	-0.612	0.305	1	0.499	0.369
λ_{++}	0.7268	0.018	-0.023	0.012	-0.048	-0.108	-0.188	-0.111	-0.107	-0.652	0.090	0.499	1	0.100
S_{ct}	0.7841	0.005	0.015	-0.000	-0.006	-0.003	-0.023	-0.720	-0.624	-0.132	0.531	0.369	0.100	1

Table 3.3: Matrix of correlation coefficients returned by the unbinned maximum likelihood fit use in the B_u lifetime measurement.

$\boldsymbol{B_s \rightarrow J/\psi(\mu^+\mu^-)\phi(K^+K^-)}$. A total of 3637 B_s candidates pass the selection applied. The results of fitting the 12-parameter model described by Equation 3.35 to this sample are collected in Table 3.4 alongside the B_u and B_d results. Fit projections onto the mass and the PDL distributions with the results of the fit quality test are shown in Figure 3-15. It should be mentioned that in calculating the PDL we use the latest $M_{B_s} = 5366.0\,\mathrm{MeV/c^2}$ from Reference [49], not the one from Reference [25].

Summary. The results of the average lifetime measurements in the three decay modes considered are summarized in Table 3.4. A number of observations are due at this point:

- Across all three samples the quality of the fit is good both in the mass and the PDL.

- There is a general similarity between the background parameters across the three samples. In fact most of them a very close. The only substantial difference is in f_{++} parameter, which is explained by reflections which are indeed different in the different samples.

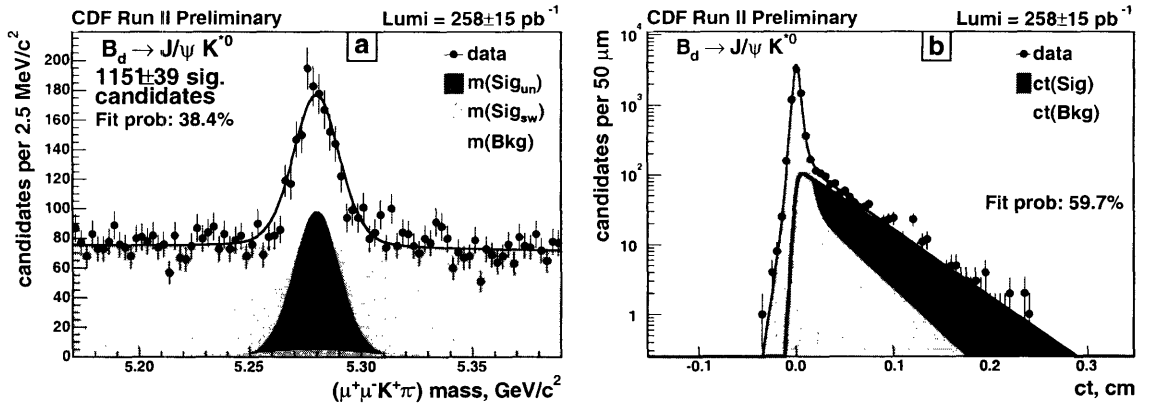


Figure 3-14: Projection of the results of the unbinned maximum likelihood fit onto (a) mass and (b) proper decay length distributions in the B_d sample.

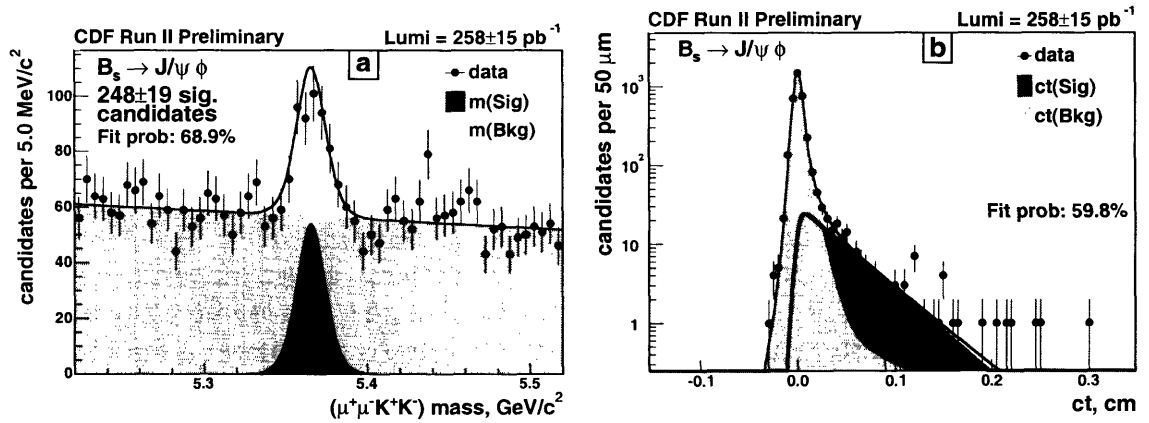


Figure 3-15: Projection of the results of the unbinned maximum likelihood fit onto (a) mass and (b) proper decay length distributions in the B_s sample.

- The uncertainty scale factors S_m and S_{ct} are compatible between B_d and B_s and are somewhat larger than those for B_u. This is expected as the B_u candidates involve combinations of only three tracks, while others are built out of four.

- τ_{B_u} and τ_{B_d} compare well to the current world averages and exhibit expected hierarchy, which confirms the validity of the lifetime extraction method used.

- The ratio $\frac{\tau_{B_u}}{\tau_{B_d}} = 1.071 \pm 0.042(stat.)$ is in agreement with the theory expectations and the previous measurements.

- τ_{B_s} being notably shorter than τ_{B_d} along with the expectation that the decay $B_s \to J/\psi\phi$ is dominated by the shorter-lived CP-even eigenstate (from the CDF Run I measurement and/or by SU(3) symmetry with $B_d \to J/\psi K^{*0}$) indicates a significant $\Delta\Gamma_s$. This motivates the time-dependent angular analysis the results of which are presented next.

Par.	B_u	B_d	B_s	Unit
M	5278.69 ± 0.26	5280.16 ± 0.40	5365.76 ± 0.73	MeV/c^2
f_s	0.1222 ± 0.0025	0.1503 ± 0.0051	0.0683 ± 0.0051	
a	-0.87 ± 0.46	-1.05 ± 0.89	-1.82 ± 0.66	(GeV/c^2)$^{-1}$
S_m	1.545 ± 0.028	1.611 ± 0.059	1.765 ± 0.108	
$c\tau$	497.4 ± 9.9	464.3 ± 15.5	408.8 ± 30.1	μm
f_-	0.0414 ± 0.0079	0.0422 ± 0.0140	0.018 ± 0.019	
f_+	0.1087 ± 0.0092	0.1455 ± 0.0195	0.1230 ± 0.0192	
f_{++}	0.0135 ± 0.0024	0.0447 ± 0.0064	0.0050 ± 0.0032	
λ_-	49 ± 4	47 ± 8	50 ± 32	μm
λ_+	55 ± 4	45 ± 6	77 ± 12	μm
λ_{++}	399 ± 60	337 ± 40	688 ± 301	μm
S_{ct}	1.245 ± 0.012	1.265 ± 0.024	1.288 ± 0.031	
N_{sig}	3333.6 ± 67.7	1150.9 ± 38.9	248.4 ± 18.6	

Table 3.4: Results of the unbinned maximum likelihood fit for the (average) lifetime in $B_u \to J/\psi K^+$, $B_d \to J/\psi K^{*0}$, and $B_s \to J/\psi\phi$ samples.

3.3.2 Time-dependent angular analysis

$B_d \to J/\psi(\mu^+\mu^-)K^{*0}(K^+\pi^-)$. A peculiarity of this decay channel is the presence of the swapped component, the handling of which is described in much detail in Section 3.2.4 and Appendices I and K. Other than this, the same 12-parameter model is used to describe the mass and the PDL of the B_d candidates as employed in the average lifetime measurement. The angular distribution of the B_d candidates is described with four parameters for signal and three parameters for background. The total number of parameters floating in the fit is 19, as the $\log \mathcal{L}$ given by Equation 3.36 implies.

The selection is unchanged from the average lifetime measurement, thus the fit is performed on the same sample of 7658 candidates. Fit results are shown in Table 3.6 and fit projections onto the data distributions are shown in Figure 3-16. For the transversity variables the projections are for the side-band subtracted and sculpting-corrected signal. The details of the procedure to make these are described in Appendix M.

$B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$, unconstrained fitting. The B_s fit allows CP-even and CP-odd angular amplitudes to have different lifetimes. In the nominal fit this is realized by using $c\tau_L$ as the lifetime of the CP-even amplitudes and $\Delta\Gamma_s/\Gamma_s$ as another fit parameter which (along with $c\tau_L$) defines the lifetime of CP-odd amplitude. Appendix N lists the relevant transformation formulae. However, as discussed in Section 1.4.2, we cannot determine $\delta_\perp$ from the B_s sample in hand, so it is not one of the fit parameters.

Compared to the average lifetime analysis the optimal cuts are different for the $\Delta\Gamma_s/\Gamma_s$ measurement, which results in a tighter and cleaner selection. Because of the tighter cuts, we have to eliminate the short negatively lived background component, (f_-, λ_-), from the fit model. Total elimination rather than fixing to some value

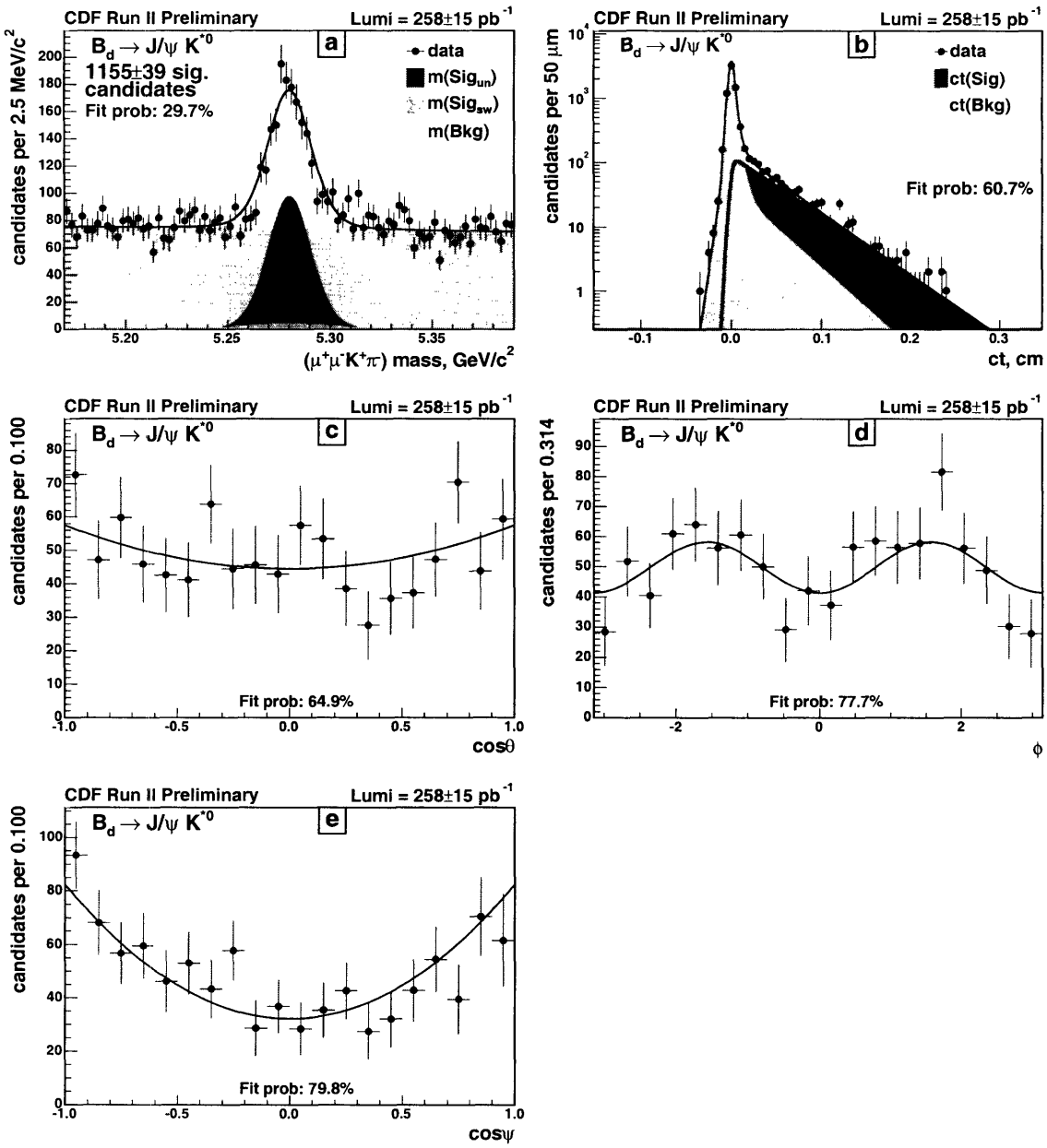


Figure 3-16: Projections of the results of the unbinned maximum likelihood fit used in time-dependent angular analysis of the $B_d \to J/\psi K^{*0}$ decay overlaid on data: (a) mass, (b) PDL, (c)† $\cos\theta$, (d)† ϕ, (e)† $\cos\psi$. [$\dagger$ – side-band subtracted and sculpting-corrected signal is plotted as explained in Appendix M]

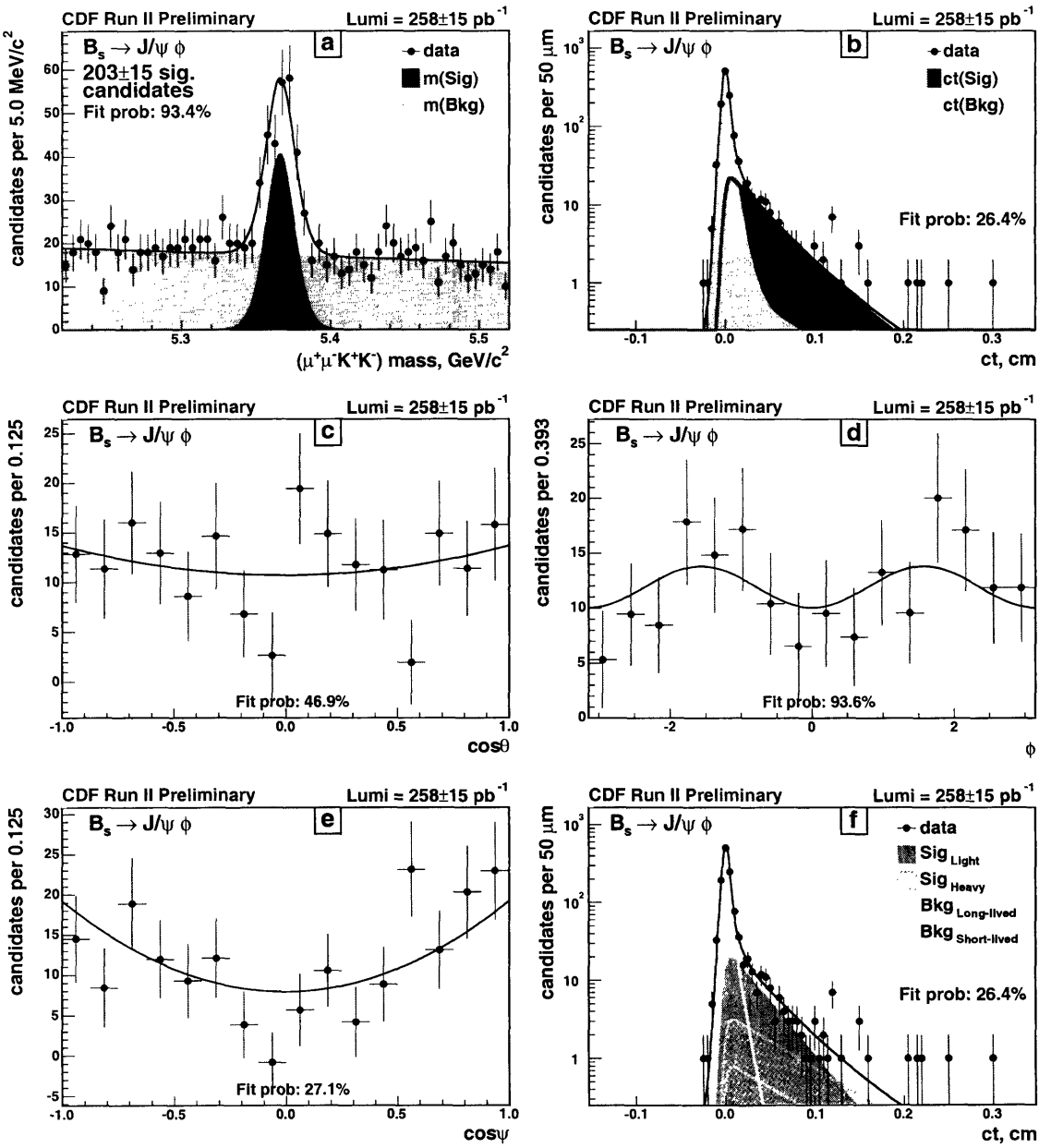


Figure 3-17: Projections of the results of the *unconstrained* unbinned maximum likelihood fit used in time-dependent angular analysis of the $B_s \to J/\psi\phi$ decay overlaid on data: (a) mass, (b) PDL, (c)† $\cos\theta$, (d)† ϕ, (e)† $\cos\psi$, (f) PDL with the contributions from light and heavy mass eigenstates shown separately. [† – side-band subtracted and sculpting-corrected signal is plotted as explained in Appendix M]

determined by a study of one kind or another, is justified if one considers the following. At relatively loose cuts used in the average lifetime measurement the significance of f_- is only about one standard deviation (Table 3.4). It has been established by fits on the optimization `NTuples` described in Appendix D that as cuts are tightened f_- becomes even less significant and eventually the fit becomes insensitive to it (does not converge). The default cuts applied in the B_s sample selection for $\Delta\Gamma_s/\Gamma_s$ analysis are even tighter than those at which the fit becomes insensitive to (f_-, λ_-) component, so it is only natural not to have this component in the fit.

The sample to which we fit our 17-parameter model (Equation 3.38 with (f_-, λ_-) eliminated) consists of 1238 candidates, both signal and background. The results of the fit are quoted in Table 3.6 alongside B_d results; the correlation matrix returned by the fit is shown in Table 3.5; and the fit projections onto the data distributions are shown in Figure 3-17.

Par.	GLOB.	M	f_s	A	S_m	$\|A_0\|^2$	$\|A_\parallel\|^2$	$\delta_\parallel$	$\|B_0\|^2$
M	0.0782	1	0.038	-0.015	0.056	-0.013	0.012	0.005	-0.000
f_s	0.2816	0.038	1	0.004	0.118	0.028	0.044	-0.025	-0.043
A	0.0157	-0.015	0.004	1	-0.001	0.001	0.000	-0.001	-0.001
S_m	0.2002	0.056	0.118	-0.001	1	-0.011	0.024	-0.026	-0.009
$\|A_0\|^2$	0.8553	-0.013	0.028	0.001	-0.011	1	-0.642	0.076	-0.105
$\|A_\parallel\|^2$	0.8860	0.012	0.044	0.000	0.024	-0.642	1	-0.207	0.053
$\delta_\parallel$	0.2429	0.005	-0.025	-0.001	-0.026	0.076	-0.207	1	0.004
$\|B_0\|^2$	0.3333	-0.000	-0.043	-0.001	-0.009	-0.105	0.053	0.004	1
$\|B_\parallel\|^2$	0.3256	0.000	0.006	0.000	-0.006	0.055	-0.091	0.017	-0.314
$arg(B_\parallel)$	0.1117	-0.004	-0.001	0.000	0.005	0.006	0.003	-0.102	-0.032
$c\tau_L$	0.7930	-0.022	-0.218	-0.003	-0.044	-0.246	0.003	0.024	0.055
f_+	0.7284	-0.017	-0.075	-0.001	0.000	-0.004	-0.007	0.008	0.007
f_{++}	0.6866	0.001	-0.031	-0.001	-0.048	0.000	0.015	-0.006	0.006
λ_+	0.7960	-0.014	-0.002	0.001	-0.073	0.029	-0.023	0.004	-0.010
λ_{++}	0.5938	0.006	0.015	-0.000	0.029	-0.002	-0.007	0.002	-0.004
S_{ct}	0.2752	-0.002	0.009	0.000	-0.005	0.002	-0.003	0.000	-0.001
$\Delta\Gamma/\Gamma$	0.8935	0.006	0.129	0.002	0.020	0.257	0.357	-0.117	-0.046

Par.	$\|B_\parallel\|^2$	$arg(B_\parallel)$	$c\tau_L$	f_+	f_{++}	λ_+	λ_{++}	S_{ct}	$\Delta\Gamma/\Gamma$
M	0.000	-0.004	-0.022	-0.017	0.001	-0.014	0.006	-0.002	0.006
f_s	0.006	-0.001	-0.218	-0.075	-0.031	-0.002	0.015	0.009	0.129
A	0.000	0.000	-0.003	-0.001	-0.001	0.001	-0.000	0.000	0.002
S_m	-0.006	0.005	-0.044	0.000	-0.048	-0.073	0.029	-0.005	0.020
$\|A_0\|^2$	0.055	0.006	-0.246	-0.004	0.000	0.029	-0.002	0.002	0.257
$\|A_\parallel\|^2$	-0.091	0.003	0.003	-0.007	0.015	-0.023	-0.007	-0.003	0.357
$\delta_\parallel$	0.017	-0.102	0.024	0.008	-0.006	0.004	0.002	0.000	-0.117
$\|B_0\|^2$	-0.314	-0.032	0.055	0.007	0.006	-0.010	-0.004	-0.001	-0.046
$\|B_\parallel\|^2$	1	-0.017	-0.009	-0.006	-0.005	0.009	0.002	0.003	-0.031
$arg(B_\parallel)$	-0.017	1	-0.004	-0.000	0.003	0.000	-0.001	-0.000	0.008
$c\tau_L$	-0.009	-0.004	1	0.065	-0.037	-0.013	0.003	-0.009	-0.690
f_+	-0.006	-0.000	0.065	1	0.154	-0.659	-0.101	-0.258	-0.031
f_{++}	-0.005	0.003	-0.037	0.154	1	-0.514	-0.579	-0.031	0.027
λ_+	0.009	0.000	-0.013	-0.659	-0.514	1	0.380	0.110	0.008
λ_{++}	0.002	-0.001	0.003	-0.101	-0.579	0.380	1	0.022	-0.022
S_{ct}	0.003	-0.000	-0.009	-0.258	-0.031	0.110	0.022	1	0.002
$\Delta\Gamma/\Gamma$	-0.031	0.008	-0.690	-0.031	0.027	0.008	-0.022	0.002	1

Table 3.5: Matrix of correlation coefficients returned by the *unconstrained* unbinned maximum likelihood fit used in the time-dependent angular analysis of the B_s sample.

We have performed `MINOS` analysis [100] of the shape of the likelihood around the fit minimum, looking if any signal parameters should be quoted with asymmetric

uncertainties. Not unexpectedly, $c\tau_L$ and $\Delta\Gamma/\Gamma$ are such two parameters. Their asymmetric uncertainties are shown in Table 3.6 instead of the parabolic ones.

$B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$, constrained fitting. Here to improve the resolution on the $\Delta\Gamma_s/\Gamma_s$ we supply the fitter with additional information, namely the theoretical prediction that Γ_s equals Γ_d to within 1% (Equation 1.86). Technically this constraint is realized by multiplying the likelihood in Equation 3.38 by a Gaussian term:

$$\mathcal{L} \longrightarrow \frac{1}{\sqrt{2\pi}\sigma_{tot}} e^{-\frac{(c\tau_{B_s} - c\tau_{B_d})^2}{2\sigma_{tot}^2}} \mathcal{L}, \tag{3.39}$$

where $c\tau_{B_s} \equiv \frac{c}{\Gamma_s} = c\tau_L(1+\frac{1}{2}\frac{\Delta\Gamma}{\Gamma})$ (see Equation N.16), $c\tau_{B_d} = 460.8\,\mu$m [25], and $\sigma_{tot} = \sqrt{\sigma_{th}^2 + \sigma_{exp}^2} = \sqrt{(10^{-2}\cdot 460.8\,\mu\text{m})^2 + (4.5\,\mu\text{m})^2} \approx 6.4\,\mu$m comes from the uncertainty in the theoretical prediction in Equation 1.86 and the experimental uncertainty of the Γ_d measurement [25]. The results of the fit with such constraint in place are collected in Table 3.6; projections of the fit results on data are shown in Figure 3-18.

Summary. The results of the nominal fits on B_d and B_s samples are placed side by side in Table 3.6. Plots (a) and (b) in Figures 3-16 – 3-18, showing the mass

Par.	B_d	B_s, unconstr. fit	B_s, constr. fit	Unit
M	5280.2 ± 0.4	5366.1 ± 0.8	5366.1 ± 0.8	MeV/c^2
f_s	0.151 ± 0.005	0.164 ± 0.012	0.162 ± 0.012	
a	-1.06 ± 0.89	-2.2 ± 1.2	-2.2 ± 1.2	(GeV/c^2)$^{-1}$
S_m	1.65 ± 0.06	1.81 ± 0.12	1.81 ± 0.12	
$\lvert A_0\rvert^2$	0.562 ± 0.025	0.615 ± 0.064	0.614 ± 0.064	
$\lvert A_\parallel\rvert^2$	0.223 ± 0.032	0.260 ± 0.086	0.291 ± 0.080	
$\delta_\parallel$	2.86 ± 0.22	1.93 ± 0.36	1.90 ± 0.32	
$\delta_\perp$	0.15 ± 0.15	—	—	
$\lvert B_0\rvert^2$	0.292 ± 0.009	0.318 ± 0.023	0.319 ± 0.023	
$\lvert B_\parallel\rvert^2$	0.358 ± 0.017	0.385 ± 0.041	0.384 ± 0.041	
$arg(B_\parallel)$	1.60 ± 0.06	1.62 ± 0.13	1.62 ± 0.13	
$c\tau_{(L)}$	461.5 ± 15.3	$316\ ^{+48}_{-40}$	$340\ ^{+40}_{-28}$	μm
$\Delta\Gamma/\Gamma$	—	$65\ ^{+25}_{-33}$	$71\ ^{+24}_{-28}$	%
f_-	0.042 ± 0.014	—	—	
f_+	0.145 ± 0.019	0.124 ± 0.031	0.126 ± 0.031	
f_{++}	0.044 ± 0.006	0.011 ± 0.007	0.011 ± 0.007	
λ_-	47 ± 7	—	—	μm
λ_+	45 ± 6	66 ± 17	66 ± 17	μm
λ_{++}	348 ± 40	634 ± 280	629 ± 278	μm
S_{ct}	1.27 ± 0.02	1.33 ± 0.04	1.33 ± 0.04	
N_{sig}	1155 ± 39	203 ± 15	201 ± 15	

Table 3.6: Results of the nominal fits on B_d and B_s samples.

and the PDL distributions for the fits and the data, are made using the respective data samples in their *entirety*. Unlike these, plots (c) – (e), are only for the *sideband*

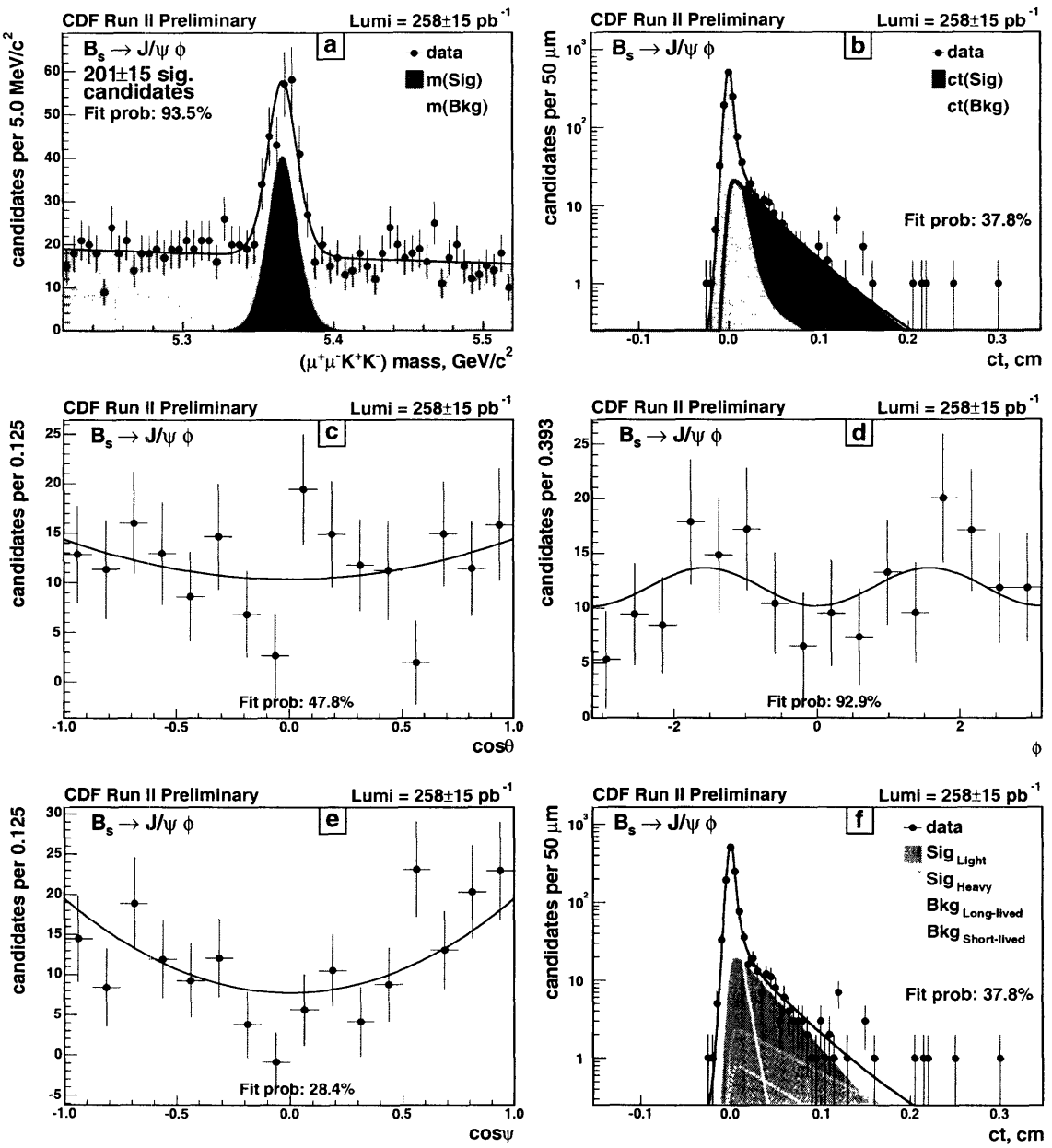


Figure 3-18: Projections of the results of the *constrained* unbinned maximum likelihood fit used in time-dependent angular analysis of the $B_s \to J/\psi\phi$ decay overlaid on data: (a) mass, (b) PDL, (c)† $\cos\theta$, (d)† ϕ, (e)† $\cos\psi$, (f) PDL with the contributions from light and heavy mass eigenstates shown separately. [† – side-band subtracted and sculpting-corrected signal is plotted as explained in Appendix M]

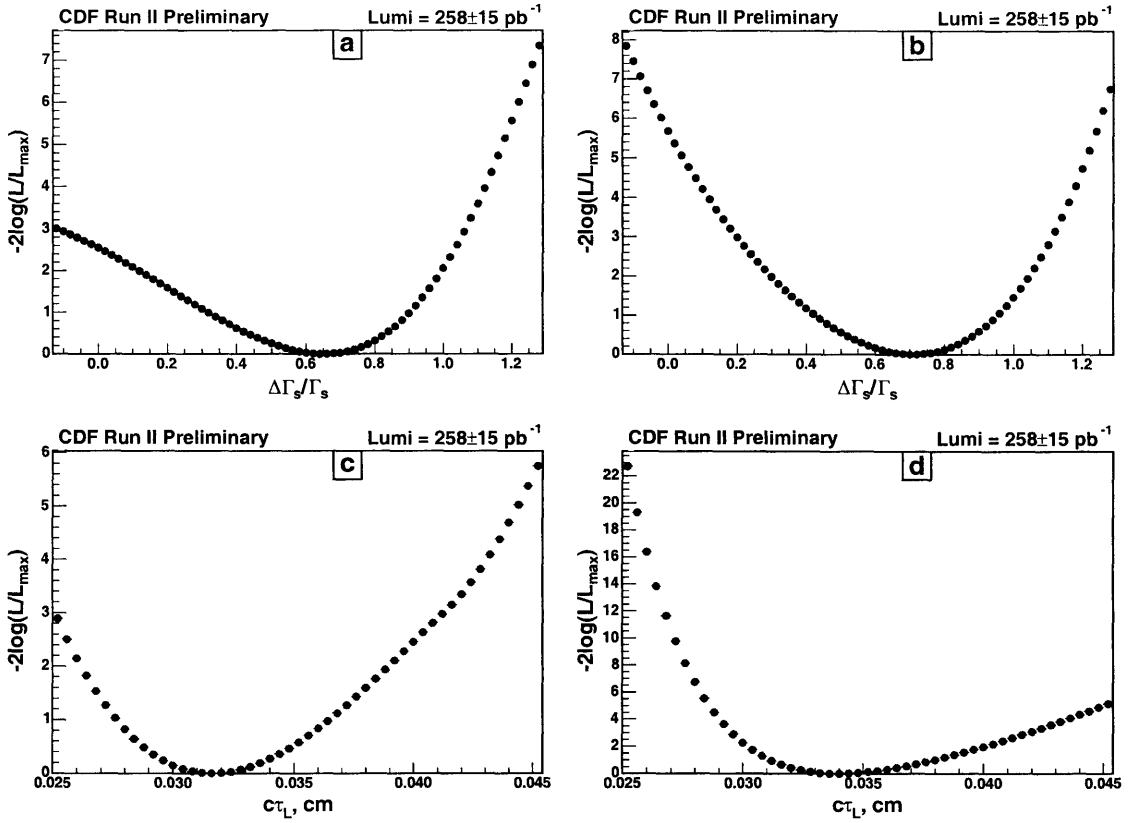


Figure 3-19: Scan of the log-likelihood as a function of a parameter of interest. To calculate each point, the parameter is fixed at a value, the fit is re-done and $-2\log(\mathcal{L}/\mathcal{L}_{max})$ is calculated ($\mathcal{L}_{max}$ is the value of the likelihood when all parameters take on optimal values). (a) $\Delta\Gamma_s/\Gamma_s$ unconstrained fit, (b) $\Delta\Gamma_s/\Gamma_s$ constrained fit, (c) $c\tau_L$ unconstrained fit, (d) $c\tau_L$ constrained fit.

subtracted signal, as explained in Appendix M. In the sideband subtraction, in order to make the errors smaller a $ct > 0$ cut is applied, which has a slight effect of enhancing the CP-odd component in B_s.

Figure 3-19 shows the scans of the log-likelihood as a function of $\Delta\Gamma_s/\Gamma_s$ and $c\tau_L$ both for unconstrained and constrained fits. These scans have a number of applications, such as:

- testing the shape at the minimum and making sure that the extracted value of a parameter does not correspond to any secondary (false) minimum,
- combining the result with other measurements,
- giving sense of how significantly away from zero or the theoretical expectation of 0.12 the measured $\Delta\Gamma_s/\Gamma_s$ is.

Finally, Table 3.7 collects numerical values for other quantities of interest in $B_s \to J/\psi\phi$, such as time-integrated fractions, C_α (Appendix J), τ_H, $\Delta\Gamma$, *etc.* useful on their

own as well as for comparisons to earlier results. Some of these quantities, especially those that have asymmetric uncertainties, are best obtained by doing the fit with different parameterization, *e.g.* with (τ_L, τ_H) instead of $(\tau_L, \Delta\Gamma/\Gamma)$ to get τ_H. Others are calculated from the results of the nominal fits using the standard procedure of error propagation including the correlations.

Par.	Unconstrained fit		Constrained fit		Unit
	Time-integrated fractions (C_α)	True amplitudes (A_α)	Time-integrated fractions (C_α)	True amplitudes (A_α)	
M_{B_s}	5366.1 ± 0.8		5366.1 ± 0.8		MeV/c^2
$\lvert A(C)_0\rvert^2$	0.549 ± 0.056	0.615 ± 0.064	0.555 ± 0.056	0.614 ± 0.064	
$\lvert A(C)_\parallel\rvert^2$	0.232 ± 0.077	0.260 ± 0.086	0.264 ± 0.071	0.291 ± 0.080	
$\lvert A(C)_\perp\rvert^2$	0.219 ± 0.076	0.125 ± 0.066	0.181 ± 0.064	0.095 ± 0.052	
$\lvert A(C)_0\rvert$	0.741 ± 0.038	0.784 ± 0.039	0.745 ± 0.037	0.783 ± 0.038	
$\lvert A(C)_\parallel\rvert$	0.481 ± 0.081	0.510 ± 0.082	0.513 ± 0.069	0.539 ± 0.070	
$\lvert A(C)_\perp\rvert$	0.469 ± 0.082	0.354 ± 0.098	0.425 ± 0.075	0.308 ± 0.087	
$\delta_\parallel$	1.93 ± 0.36		1.90 ± 0.32		
$c\tau_L$	$316\ ^{+48}_{-40}$		$340\ ^{+40}_{-28}$		μm
$c\tau_H$	$622\ ^{+175}_{-138}$		$713\ ^{+167}_{-129}$		μm
$c\tau_s$	$419\ ^{+45}_{-38}$		**460.8 ± 6.4** (constrained)		μm
$\Delta\Gamma_s/\Gamma_s$	$65\ ^{+25}_{-33}$		$71\ ^{+24}_{-28}$		%
$\Delta\Gamma_s$	$0.47\ ^{+0.19}_{-0.24}$		0.46 ± 0.18		ps^{-1}
N_{sig}	203 ± 15		201 ± 15		

Table 3.7: Summary of the signal parameters obtained with unconstrained and constrained fitting of the B_s sample with various parameterizations. Asymmetric uncertainties are quoted where deemed more appropriate than parabolic.

3.4 Systematic uncertainties

As almost always is the case, our measurements are, at least potentially, influenced by biases and systematic effects. A correct assessment of these is an essential part of the analyses. The current section is a collection of sub-sections, each one discussing a particular systematic effect and its influence on the average lifetime measurement and/or time-dependent angular analysis.

3.4.1 Selection

There is no reason to believe that the cuts used in selecting the B candidates can bias the lifetime measurements. It is however desirable to confirm that it is indeed is the case. Figure 3-20 shows (in)dependence on the extracted B_u lifetime of the p_T cuts used in the selection. The slightly shorter lifetimes at very loose cuts have to do with the badly mis-reconstructed candidates making their into the sample, *i.e.* the inadequacy of the model. This is accounted for in Section 3.4.3, where the systematic

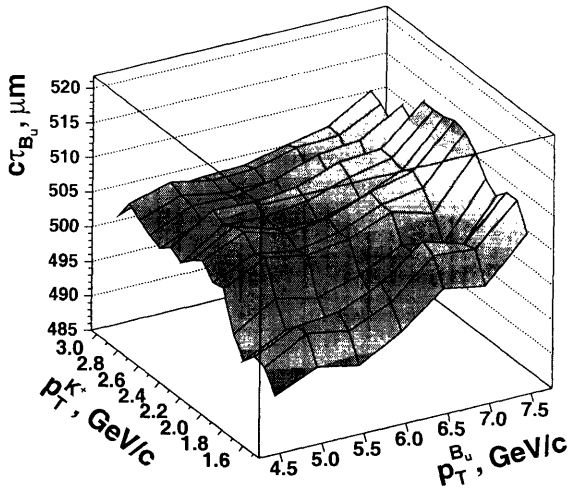


Figure 3-20: Lifetime extracted from the sample of $B_u \to J/\psi K^+$ decays. All other cuts are as defined in Table 3.1.

uncertainty due to the choice fit model is evaluated. We also include (Figure 3-21) plots showing B_u lifetime extracted at different values of one of the p_T cuts while the other one is kept fixed. We do not assign any systematic uncertainty due to particular choice of p_T cuts, which is in line with statisticians' advice on practical matters such as this one [104].

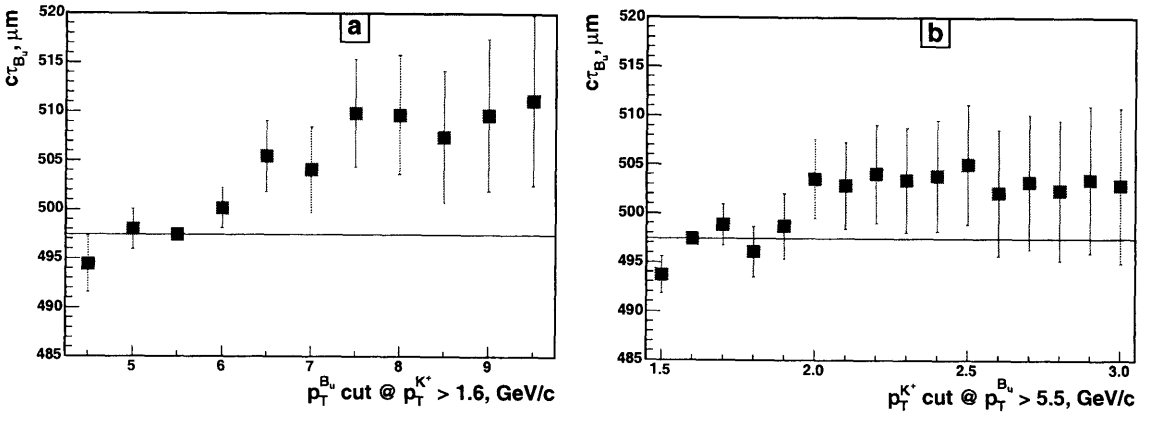


Figure 3-21: B_u lifetime extracted at the different values of (a) $p_T^{B_u}$ (b) $p_T^{K^+}$ cut. Points with no error bars correspond to the nominal cuts. Error bars on all other points represent an estimate of a probable shift in the lifetime due to the change in the sample when going from the nominal cuts to the cuts represented by the point in question.

The only other non-technical cut applied is a vertex quality cut, which we apply in the form $Prob(\chi^2) > \epsilon$. B_u lifetimes extracted at different values of this cut are shown in Figure 3-22. Again, there are no variations incompatible with statistical

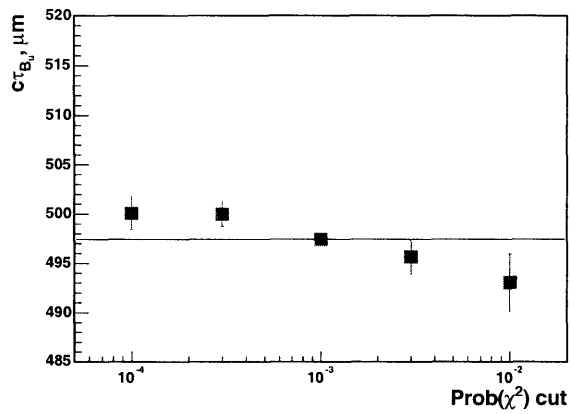


Figure 3-22: B_u lifetime extracted at the different values of the vertex quality, $Prob(\chi^2)$, cut. Points with no error bars correspond to the nominal cuts. Error bars on all other points represent an estimate of a probable shift in the lifetime due to the change in the sample when going from the nominal cut to the cut represented by the point in question.

change in the sample when going from one value of the cut to another.

In what concerns the amplitudes extracted in the time-dependent angular analysis, the selection could hardly have an effect. Indeed, we do know that the angular distributions are sculpted by the cuts, as, for example, is illustrated by Figure 3-8. However, this sculpting is taken into account as long as the Monte-Carlo simulation used in calculating ξ_i is accurate. The effect of any residual disagreement between distributions produced by Monte-Carlo and the respective ones from data is discussed in Section 3.4.4.

3.4.2 SVX alignment

A history of the average B meson lifetime measurements at CDF during Run II using successively larger samples has a pattern of the biggest systematic uncertainty stemming from the imperfect knowledge of positions of the SVX wafers, also known as SVX alignment limitations. Over time the precision of the alignment has improved, and it is now reasonable to assume that no major improvements are possible. In other words, the alignment seems to be stable and of high quality. A way to estimate SVX alignment effects on the extracted lifetime is to vary (distort) the alignment in some reasonable fashions and see what these variations do to the extracted lifetime. We determined [101] that the following four alignment distortions are suitable for estimating the systematic uncertainty:

1. 50 μm move-out – all ladders are universally shifted away from the central axis (in a radial direction) by 50 μm (`1x0047 1 TEST`)[2].

[2]The label in parentheses gives a CDF-specific code for the particular alignment table. "x" in these codes can be 4, 5, or 6 and refers to different periods of SVX mechanical configuration being stable.

2. 50 μm move-in – all ladders are universally shifted toward the central axis by 50 μm (`1x0047 2 TEST`).

3. 50 μm bow-out – in each ladder the two central wafers (the total number of wafers in a ladder is four) are universally shifted away from the central axis by 50 μm (`1x0047 3 TEST`).

4. 50 μm bow-in – in each ladder the two central wafers are universally shifted toward the central axis by 50 μm (`1x0047 4 TEST`).

We generate 10M $B_u \to J/\psi K^+$ decays, which are then subjected to the realistic detector and trigger simulation. Events that pass the trigger are subsequently run through production. The resulting simulated data sample is reconstructed using the software identical to the actual analysis software. In the reconstruction tracks are refitted using the same alignment table that was used in the simulation/production as well as using the four distorted alignments described above. After typical selection we end up with about 240K analysis quality signal B_u candidates, on which the lifetime fit is performed: a single exponential convoluted with a Gaussian of event-by-event width is fitted to the PDL distribution. The extracted fit parameters (lifetime and the PDL uncertainty scale factor) are tabulated in Table 3.8.

Alignment	Distortion	$c\tau$, μm	$\Delta c\tau$, μm	S_{ct}	$N_{B\ cand.}$
`160047 1 TEST`	move-out	498.60 ± 1.02	$+1.22$	1.133 ± 0.009	240404
`160047 2 TEST`	move-in	496.07 ± 1.01	-1.31	1.119 ± 0.009	241104
`160047 3 TEST`	bow-out	498.02 ± 1.02	$+0.64$	1.112 ± 0.009	241032
`160047 4 TEST`	bow-in	496.74 ± 1.01	-0.64	1.109 ± 0.009	241465
original	none	497.38 ± 1.01	0.00	1.084 ± 0.009	242610

Table 3.8: Results of the lifetime fit on the signal Monte-Carlo sample of $B_u \to J/\psi K^+$ decays, reconstructed using different SVX alignments.

In Table 3.8 we see more or less what we expect to see, namely:

- We lose some candidates with distorted alignments compared to when the original one is used.

- Shifts are bigger distortions than bows (four wafers are moved, not just the central two), thus bigger effects for shifts than for bows.

- For larger distortions the uncertainty scale factor is larger and for the undistorted alignment it is the smallest.

- The "in-" distortions bias extracted lifetime in the opposite direction to the "out-" distortions.

Comparing the result obtained with the original alignment to those we get with distorted ones, we assign a systematic uncertainty due to SVX alignment imperfections of ± 1 μm.

A lesser effect is expected from potential mis-alignment in z direction. Indeed, z mis-alignment can only affect lifetime measurements indirectly, perhaps through candidate selection or large track parameter correlations. To check if there is any substantial effect, we compare the lifetime results obtained with the two calibration tables (`PASS 11A` and `PASS 11`), one of which includes the nominal alignment and the other one has an alignment with the SVX barrel 3 somewhat misplaced along the z axis. We observe what we expect to observe: with a worse alignment we reconstruct 3.6 fewer signal candidates, the error scale factor is worse by 0.0018, *etc.* The change in the lifetime is 0.45 μm, which should be covered by the above assignment of ± 1 μm as a systematic uncertainty due to imperfect SVX alignment.

The alignment effects on the angular decay amplitudes do not appear to be significant.

3.4.3 Fit model

In this sub-section we consider systematic uncertainties that arise when we make choices about the fit model. Some of the parameterizations we choose cannot be derived from basic principles, at least not in a very straightforward manner. We regard the following few as pertaining to this category:

- ct resolution function: what should it be – single Gaussian, double Gaussian, Gaussian with exponential tails?
- Background ct parameterization: do we give the fitter not enough or too much freedom?
- Background angular distribution parameterization.

Below we attempt to estimate how much the extracted parameters would have differed had we made other parameterization choices.

***ct* resolution function**. We have chosen a single Gaussian as the PDL resolution function. This choice was made by looking at the ct distribution of three and four prompt track combinations. However, we do not have a clean and large enough sample of vertices *guaranteed* to be prompt. There are various mis-reconstructions as well as long-lived component, which limit what we can do to establish the exact shape of the resolution function. A single Gaussian is a good approximation, but we need to study other similar shapes to see how big of an effect on the extracted B lifetime the choice of a particular one of them has.

Reasonable variations of the resolution function include a double Gaussian and a Gaussian with exponential tails. We first discuss the latter option.

In the earlier analysis we did a successful fit to data using a single Gaussian with exponential tails as the PDL resolution function [102]. The tails of the resolution function were determined to be only $1.3 \pm 0.9\%$ of the Gaussian core, *i.e.* not statistically significant. Tracks in the current analysis are of superior quality due to improved tracking algorithm, better SVX alignment and usage of the SVX z and small-angle stereo hits, which leads us to expect even smaller tails on the resolution function.

Indeed, it seems impossible to get the fit to converge when the characteristic length of the tails is let to float. If we fix it to 320 μm [102] we get a fraction compatible with zero within 0.5σ and no appreciable change in the extracted B lifetime. This tells us that a Gaussian with exponential tails is not a viable option for the resolution function, or, more precisely, it degenerates into a single Gaussian.

Next we try a double Gaussian as a resolution function. It turns out to be exceedingly hard to get the fit to converge without fixing some of the parameters. This tells us that double Gaussian is not a likely candidate for the resolution function shape either.

We actually did work to make this the case. Indeed, in the selection of B candidates (Section 3.1.4, Appendix F) we have excluded z regions that correspond to SVX bulkheads. Largely the motivation to do so has been to reduce the number of badly mis-reconstructed candidates. However, even if reconstructed correctly, the candidates from these regions could have ct uncertainties over-/under-estimated to a different extent with respect to the rest of the candidates, leading to a second component in the resolution function. We got rid of this complication as a by-product.

As another indication that a double Gaussian is not a viable description of the resolution function we have plotted the residuals (Figure 3-23) from projecting the results of the nominal B_u lifetime fit onto the PDL distribution. The inlays in Figure 3-23 show the residuals obtained in Reference [103] when the core of the transverse flight distance distribution in the inclusive J/ψ sample[3] is fit with a single Gaussian (left) and the double Gaussian (right). As can be deduced from these plots, the residuals should have two large, *i.e.* statistically significant, symmetrical dips around $ct = 0$ if a single Gaussian resolution function is fit where two or more Gaussians are needed. We do not observe such behavior in our case of the fully reconstructed B mesons; the distribution of residuals is compatible with flat zero line, confirming that a single Gaussian is an adequate description of the resolution function.

It is ultimately the limited statistics that does not allow us to discern the shape of the resolution function from a single Gaussian. On the other hand, it is just a different way of saying that the resolution function *is* a single Gaussian. Table 3.9 summarizes the fit results with different parameterizations of the resolution function.

Background *ct* modeling. In this category we consider two modifications of the fit model that affect model's flexibility. As the first alteration (Model A), we add another short positively lived exponential, whose shape and size are fixed to those of the (f_-, λ_-) component. Thus the change in notation from (f_-, λ_-) to (f_{-+}, λ_{-+}). In principle, this makes the model more flexible than the one we use by default. With the second alteration (Model B) we attempt to make the model less flexible. We do exactly what we did with the first alteration, except that on the positive ct side we allow only one independently varied exponential for the background, not two that the normally have.

[3]The resolution function in that case appears to be a multi- or, at the very least, a double-Gaussian.

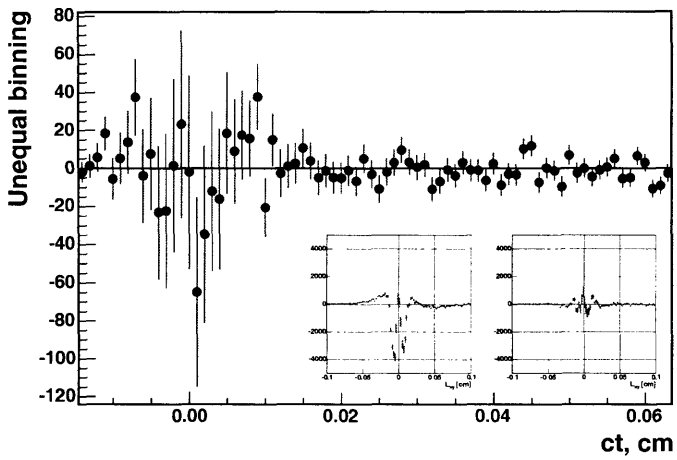


Figure 3-23: Residuals from projecting the results of the fit onto data. Before the residuals are calculated the data are binned such that there are at least 20 entries in each bin while bin width is a multiple of 10 μm. A case of single (left) and double (right) Gaussian fit to the transverse flight distance of inclusive J/ψ is shown on the inlays.

Par.	Nominal model	Res. f-n: $G_1 + G_2$	Res. f-n: $G + ET$	Units
M	5278.94 ± 0.27	5278.95 ± 0.27	5278.94 ± 0.27	MeV/c^2
f_s	0.1840 ± 0.0037	0.1848 ± 0.0037	0.1840 ± 0.0037	
A	-1.53 ± 0.62	-1.53 ± 0.62	-1.53 ± 0.62	$(\text{GeV/c}^2)^{-1}$
S_m	1.534 ± 0.029	1.537 ± 0.029	1.534 ± 0.029	
$c\tau$	505.78 ± 10.35	504.11 ± 10.30	505.75 ± 10.35	μm
f_-	0.0323 ± 0.0061	**0.03** (fixed)	0.339 ± 0.0066	
f_+	0.1137 ± 0.0098	0.0850 ± 0.0099	0.1147 ± 0.0096	
f_{++}	0.0198 ± 0.0039	0.0203 ± 0.0039	0.0197 ± 0.0033	
λ_-	72 ± 8	**300** (fixed)	69 ± 10	μm
λ_+	63 ± 6	58 ± 8	62 ± 4	μm
λ_{++}	393 ± 59	386 ± 58	393 ± 54	μm
S_{ct}	1.278 ± 0.015	1.256 ± 0.017	1.276 ± 0.015	
		f^{G_2} : 0.069 ± 0.013 $S_{ct}^{G_2}$: 3.35 ± 0.24	f^{ET}: $(4.1 \pm 9.5)10^{-4}$ λ^{ET}: **320 μm** (fixed)	
N_{sig}	3012.9 ± 60.3	3026.0 ± 60.4	3013.3 ± 60.3	

Table 3.9: Results of the unbinned maximum likelihood fit on the $B_u \to J/\psi K^+$ sample using the nominal model and the variations described in the text. In getting these numbers we use a tighter set of cuts, $p_T^{B_u} > 6.0\,\text{GeV/c}$, $p_T^{K^+} > 2.0\,\text{GeV/c}$, to be even more certain that precarious badly mis-reconstructed events do not affect our judgment.

The results of the fits with these two models are collected in Table 3.10. The lifetime extracted from the B_u sample with Model A comes out to be only 0.03 μm shorter than the one the default model gives. Fit with Model B yields a lifetime that is 3.5 μm longer than the nominal one. It is worth mentioning however, that this last model is a deviation from the nominal one that is somewhat beyond the borderline of being reasonable, as the discussed below fit of the PDL of B_u mass side-bands suggests.

Par.	Nominal model	Model A	Model B	Units
M	5278.69 ± 0.26	5278.69 ± 0.26	5278.62 ± 0.26	MeV/c^2
f_s	0.1222 ± 0.0025	0.1222 ± 0.0025	0.1242 ± 0.0025	
A	-0.87 ± 0.46	-0.87 ± 0.46	-0.91 ± 0.46	(GeV/c^2)$^{-1}$
S_m	1.545 ± 0.028	1.545 ± 0.029	1.568 ± 0.028	
$c\tau$	497.4 ± 9.9	497.4 ± 9.9	500.9 ± 9.7	μm
$f_{-(+)}$	0.0414 ± 0.0079	0.0417 ± 0.0085	0.0667 ± 0.0119	
f_+	0.1087 ± 0.0092	0.0676 ± 0.0076	0.0411 ± 0.0041	
f_{++}	0.0135 ± 0.0024	0.0132 ± 0.0025	—	
$\lambda_{-(+)}$	49 ± 4	49 ± 5	41 ± 3	μm
λ_+	55 ± 4	59 ± 7	160 ± 15	μm
λ_{++}	399 ± 60	404 ± 62	—	μm
S_{ct}	1.245 ± 0.012	1.245 ± 0.013	1.240 ± 0.014	
N_{sig}	3333.6 ± 67.7	3333.6 ± 67.7	3390.1 ± 67.4	

Table 3.10: Results of the unbinned maximum likelihood fit on the $B_u \to J/\psi K^+$ sample using the nominal model and two models with altered background parametrization.

In modeling the PDL of the background we assume, as the similarity in PDL distributions of the left and right mass side-bands suggests (Figure 3-6 (b)), that the nominal background PDL model works reliably in all regions of the mass distribution. In order to verify this, we fit the shape given by Equation 3.7 to the PDL of the candidates in the side-bands. We then fix the parameters of the nominal model to the values obtained in this fit and perform the fit on the entire B_u sample. The results of this exercise are collected in Table 3.11. When fixing the background PDL parameterization to that obtained from the fit to the side-bands the extracted lifetime shifts down by 3.2 μm. This, however, is an overestimate of any potential effect, for at least two reasons:

1. The side-bands are not completely signal free; there is $B_u \to J/\psi\pi^+$ contribution in the right side-band.

2. The side-bands include only about half of all background, therefore any parameter measured in just the sidebands, *i.e.* with less statistics than available, is likely to deviate from the true value by more than the same parameter measured with full statistics.

Par.	Nominal fit	Side-bands only	Bkg. fixed from side-bands	Units
M	5278.69 ± 0.26	—	5278.74 ± 0.26	$\mathrm{MeV/c^2}$
f_s	0.1222 ± 0.0025	—	0.1198 ± 0.0024	
A	-0.87 ± 0.46	-0.38 ± 0.48	**-0.38** (fixed)	$(\mathrm{GeV/c^2})^{-1}$
S_m	1.545 ± 0.028	—	1.515 ± 0.027	
$c\tau$	497.4 ± 9.9	—	494.2 ± 9.7	μm
f_-	0.0414 ± 0.0079	0.0340 ± 0.0080	**0.0340** (fixed)	
f_+	0.1087 ± 0.0092	0.0920 ± 0.0095	**0.0920** (fixed)	
f_{++}	0.0135 ± 0.0024	0.0174 ± 0.0024	**0.0174** (fixed)	
λ_-	49 ± 4	54 ± 6	**54** (fixed)	μm
λ_+	55 ± 4	64 ± 6	**64** (fixed)	μm
λ_{++}	399 ± 60	469 ± 48	**469** (fixed)	μm
S_{ct}	1.245 ± 0.012	1.260 ± 0.015	**1.260** (fixed)	
N_{sig}	3333.6 ± 67.7	—	3268.8 ± 64.2	

Table 3.11: Results of the unbinned maximum likelihood fit to the mass side-bands in the $B_u \to J/\psi K^+$ sample, as well as the fit to the entire mass region using the parameters obtained from the side-band-only fit.

Lastly, in support of the claim made two paragraphs above, we would like to note that the fit to the side-bands is robust in discerning between (f_+, λ_+) and (f_{++}, λ_{++}) using just the PDL information, which establishes that there are indeed at least two positively lived exponentials required in the background PDL description.

Other *ct* modeling issues. Now we consider the systematic uncertainty due to an admixture of badly mis-reconstructed candidates which are not explicitly parameterized in our nominal model. We have touched upon this in the section on cut optimization and now we have to evaluate the effect quantitatively.

With some loose cuts we fit an exponential to the PDL distribution in the range from $-2000\,\mu$m to $-400\,\mu$m (Figure 3-24). We then extrapolate the fitted line to and beyond $ct = 0$ (assuming a shape symmetric around $ct = 0$) and estimate a fraction of background that is described in such a way, as well as the characteristic decay length of these candidates. The two numbers come out to be about 0.13% and about $350\,\mu$m. We then add such a contribution to the model and perform a fit on the sample. The results of the fit with the such constructed model are compared to those obtained with the nominal model in Table 3.12. It is interesting to see that there are some noticeable changes in parameters describing the PDL of the background, however the signal lifetime changes by only $0.26\,\mu$m.

The signal could also have some amount of mis-reconstruction, but judging from what we see in the background the amount of signal mis-reconstruction is expected to be tiny and play no role in the fit.

***ct* modeling summary**. The largest changes in the extracted lifetime we have seen when tweaking the model are about $\pm 3.4\,\mu$m. They are due to somewhat unreasonable changes in the model. The second largest change is $-1.7\,\mu$m again due to less than satisfying choice of a double Gaussian for the resolution function. The

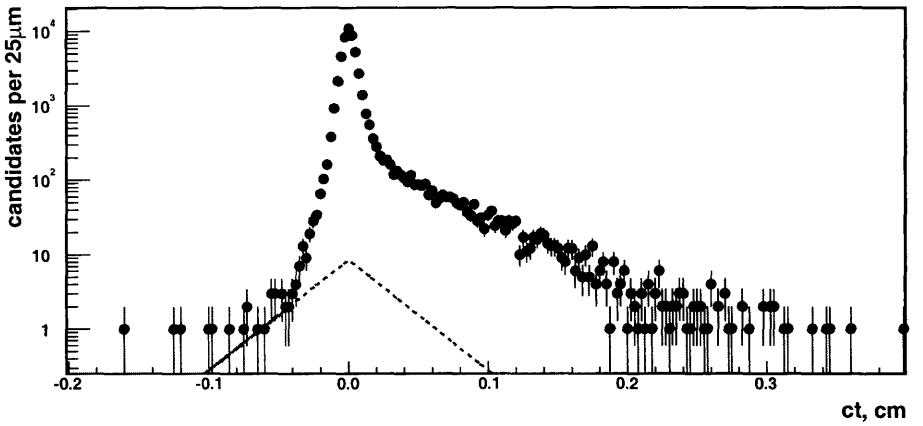


Figure 3-24: Illustration for the estimation of the systematic effect of the badly mis-reconstructed candidates on the extracted average lifetime. B_u sample is used.

Par.	Nominal	Nominal + mis-reco	Units
M	5278.67 ± 0.26	5278.67 ± 0.26	MeV/c^2
f_s	0.1223 ± 0.0025	0.1223 ± 0.0025	
A	-0.85 ± 0.46	-0.85 ± 0.46	$(\text{GeV/c}^2)^{-1}$
S_m	1.544 ± 0.028	1.544 ± 0.028	
$c\tau$	498.1 ± 9.9	498.4 ± 9.9	μm
f_-	0.0373 ± 0.0071	0.0520 ± 0.0117	
f_+	0.1057 ± 0.0087	0.1137 ± 0.0103	
f_{++}	0.0135 ± 0.0025	0.0127 ± 0.0025	
λ_-	52 ± 5	40 ± 4	μm
λ_+	57 ± 5	54 ± 5	μm
λ_{++}	397 ± 59	390 ± 64	μm
S_{ct}	1.249 ± 0.012	1.237 ± 0.013	
N_{sig}	3385.6 ± 68.2	3384.9 ± 68.3	

Table 3.12: Comparison of the fit results using the nominal model and the model that includes a contribution of badly mis-reconstructed events ($B_u \to J/\psi K^+$).

rest of the attempted changes in the model result in even smaller changes in the extracted lifetime. With these considerations we believe it makes sense to assign $\pm 1.7\,\mu$m systematic uncertainty due to the choice of the fit model.

Modeling the angular distribution of the background. The dominant part of the background candidates in our samples are combinations of prompt tracks. It is unclear if there should be any angular correlation among these tracks, except that two of them are likely to originate from a true $J/\psi \to \mu^+\mu^-$ decay. It is further unclear how the "true" background angular distribution is sculpted. We assume that a shape similar to the angular distribution of the signal, but with different amplitudes, describes the background. As pointed out in the summary of Section 3.3.2 our ability to check how well the assumption works by projecting results of the fit on data is

B_d^0	Nominal model	Flat background	Syst. uncert. assigned
$c\tau$, μm	461.5 ± 15.3	465.6 ± 15.5	± 3.9
$\|A_0\|^2$	0.562 ± 0.025	0.549 ± 0.025	± 0.013
$\|A_{\|\|}\|^2$	0.223 ± 0.032	0.229 ± 0.032	± 0.006
$\delta_{\|\|}$	2.86 ± 0.22	2.87 ± 0.22	± 0.01
δ_-	0.15 ± 0.15	0.16 ± 0.15	± 0.01

Table 3.13: Comparison of the B_d fit results using two different background models.

B_s^0	Nominal model	Flat background	Syst. uncert. assigned
$c\tau_L$, μm	316.3 ± 43.2	320.0 ± 43.6	± 3.7
$\Delta\Gamma/\Gamma$	0.652 ± 0.284	0.645 ± 0.291	± 0.007
$\|A_0\|^2$	0.615 ± 0.064	0.604 ± 0.061	± 0.011
$\|A_{\|\|}\|^2$	0.260 ± 0.086	0.274 ± 0.085	± 0.013
$\delta_{\|\|}$	1.93 ± 0.36	1.97 ± 0.36	± 0.03

Table 3.14: Comparison of the B_s fit results using two different background models.

limited by lack of knowledge of $V^{bkg}(\vec{\kappa})\epsilon^{bkg}(\vec{\omega},\vec{\kappa})$. It is reasonable to assume however, that the shift in the extracted quantities of interest due to imperfect modeling of the background angular distribution is not larger than the observed shift in these quantities when we replace our nominal background model with flat distribution. Thus we ascribe the difference of the results between the nominal fit and the fit with flat angular background distribution as a systematic uncertainty due to background angular parameterization. The relevant numbers are collected in Tables 3.13 and 3.14.

3.4.4 Residual disagreement between data and Monte-Carlo

As mentioned in the earlier sections, in treating the sculpting of the angular distributions one needs to calculate the partial normalization constants using an appropriate Monte-Carlo sample. This calculation depends upon the Monte-Carlo simulation accurately describing the data. Appendix C discusses this issue and demonstrates an agreement between the Monte-Carlo and the data in several distributions of various quantities of interest.

To quantify the effect of any residual difference between the Monte-Carlo and the data, we proceed as follows:

1. Identify the kinematic variables which have the largest discrepancies, using B_d data and the Monte-Carlo samples.

2. Re-weight the events in the Monte-Carlo sample so as to force the data and the Monte-Carlo histograms to be identical.

3. Use the re-weighted sample to compute a new set of partial normalization constants.

4. Apply the newly obtained partial normalization constants in the B_d fitter.

5. Obtain the systematic uncertainties for the measurements in time-dependent angular analysis of $B_d \to J/\psi K^{*0}$ by comparing the new fit results with the nominal fit results.

6. Assign the same systematic uncertainties for the corresponding quantities measured in the $B_s \to J/\psi\phi$ sample.

In this study $p_T^{K^{*0}}$ and $\Delta\phi(\mu)$ distributions (Figure C-2) are the most compelling to re-weight. Having re-weighted these distributions and re-calculated the ξ_i's we observe negligible shifts in the extracted values compared to other systematic effects and statistical uncertainty. Hence we do not assign any systematic uncertainty due to the residual disagreement between the data and the Monte-Carlo simulation in handling the sculpting of the angular distributions.

3.4.5 B_d candidates with K $\leftrightarrow$ π mass mis-assignment

Our procedure of treating B_d candidates with K $\leftrightarrow$ π mass mis-assignment described in Section 3.2.4 might potentially introduce a bias to the extracted amplitudes. The two sources of bias that we consider are as follows:

- neglecting the dependence of the ratio of the efficiencies, $R(\vec{\omega}, \vec{\kappa})$, for swapped and unswapped components on "the rest of the kinematic variables" $\vec{\kappa}$,

- a dependence of the total swapped fraction on the actual amplitudes, *e.g.* in the phase-space Monte-Carlo sample we observe that about 10.5% are reconstructed as swaps, while in the sample generated with realistic amplitudes only about 9.5% are.

In order to see how big these effects are we test the procedure on the Monte-Carlo sample generated with amplitudes from Reference [56]. This sample is discussed in some more details in Appendix C. Table 3.15 summarizes the results we get on five non-overlapping sub-samples as well as on the full sample. As one can see, there is a small residual disagreement, most notably in $|A_0|^2$. To assign the systematic uncertainties we take the observed deviations of the fit results on the entire sample from the input values. Thus assigned uncertainties are compiled in Table 3.16.

3.4.6 Residual $B_d \leftrightarrow B_s$ cross-feed

Another systematic effect which needs to be looked at is the cross-feed between the two decay channels we consider in the time-dependent angular analysis. Indeed, when a π track from the K^{*0} is assigned the K mass it may end up in the ϕ mass window which then can lead to an erroneous B_s candidate. The reverse (assigning the π mass to a K from ϕ) can lead to a real B_s faking a B_d.

This issue is discussed in Appendix E mostly from the point of view of making a selection that minimizes the cross-feed, while keeping the bulk of the respective

Par.	Input	Sub-sample A (25K)	Sub-sample B (28K)	Sub-sample C (27K)
$\|A_0\|^2$	0.597	$+0.5897 \pm 0.0047$	$+0.5940 \pm 0.0044$	$+0.5904 \pm 0.0045$
$\|A_\|\|^2$	0.243	$+0.2513 \pm 0.0069$	$+0.2464 \pm 0.0065$	$+0.2388 \pm 0.0066$
$\delta_\|$	2.5	$+2.5466 \pm 0.0412$	$+2.5207 \pm 0.0372$	$+2.5487 \pm 0.0385$
$\delta_\perp$	-0.17	-0.1262 ± 0.0296	-0.2046 ± 0.0279	-0.1731 ± 0.0281

Par.	Input	Sub-sample D (26K)	Sub-sample E (26K)	Total (134K)
$\|A_0\|^2$	0.597	$+0.5913 \pm 0.0046$	$+0.5907 \pm 0.0046$	$+0.5912 \pm 0.0020$
$\|A_\|\|^2$	0.243	$+0.2445 \pm 0.0065$	$+0.2540 \pm 0.0067$	$+0.2469 \pm 0.0020$
$\delta_\|$	2.5	$+2.5414 \pm 0.0367$	$+2.5240 \pm 0.0374$	$+2.5356 \pm 0.0170$
$\delta_\perp$	-0.17	-0.1834 ± 0.0282	-0.1788 ± 0.0292	-0.1743 ± 0.0127

Table 3.15: Results of the fit on the Monte-Carlo signal samples generated with amplitudes close to those observed in Nature.

Par.	$\|A_0\|^2$	$\|A_\|\|^2$	$\|A_\perp\|^2$	$\delta_\|$	$\delta_\perp$
Syst. uncert. assigned	± 0.006	± 0.004	± 0.002	± 0.04	—

Table 3.16: Systematic uncertainties assigned due to the way swapped signal component is handled in the time-dependent analysis of $B_d \to J/\psi K^{*0}$.

signals. Appendix E also provides an estimate of how much of the residual cross-feed there is. A concise summary of this estimate: with the analysis cuts we use the B_s contamination in the B_d sample is extremely small, while the B_d contamination of the B_s sample, at approximately 3.5%, is about an order of magnitude bigger.

In the current section we evaluate how such B_d contamination distorts our measurements in the B_s sample. The necessary Monte-Carlo samples are:

- 15M (generated) $B_d \to J/\psi(\mu^+\mu^-)K^{*0}(K^+\pi^-)$ with realistic amplitudes (Sample 6 from Appendix C),
- 10M (generated, 7.65M CP-even + 2.35M CP-odd) $B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ (Samples 4 and 5 described in Appendix C).

To evaluate the systematic effect due to the unwanted B_d admixture we first fit the pure B_s sample and then repeat the fit after adding an appropriate amount of B_d to the B_s sample. The "appropriate amount" is determined simply from the ratio of production fractions and the decay branching fractions (the difference between the reconstruction efficiencies is taken care of via the realistic detector simulation):

$$\frac{N(B_d)}{N(B_s)} = \frac{f_d}{f_s} \cdot \frac{Br(B_d \to J/\psi K^{*0}) \cdot Br(K^{*0} \to K^+\pi^-)}{Br(B_s \to J/\psi\phi) \cdot Br(\phi \to K^+K^-)} \approx 7, \tag{3.40}$$

which means we should take about 7 times fewer *generated* B_s decays than *generated* B_d decays. We do the fit on 2.1M (generated) B_s events (with appropriate mixture of CP-even and CP-odd components) with and without contamination from 15M (gen-

Par.	Pure B_s	B_s with B_d admixture	Syst. uncert. assigned
$\|A_0\|^2$	0.5388 ± 0.0043	0.5387 ± 0.0043	—
$\|A_{\|\|}\|^2$	0.2034 ± 0.0065	0.2064 ± 0.0064	± 0.003
$\delta_{\|\|}$	1.9479 ± 0.0315	1.9477 ± 0.0307	—
$c\tau_L$, μm	333.9 ± 3.1	338.9 ± 3.1	-5.0
$\Delta\Gamma_s/\Gamma_s$	0.4812 ± 0.0203	0.4734 ± 0.0202	± 0.008
N_{sig}	26694	27701	

Table 3.17: Comparison of the fit results on the B_s signal Monte-Carlo sample with and without B_d contamination.

erated) B_d events. The results of these fits and the assigned systematic uncertainties are summarized in Table 3.17.

3.4.7 Unequal number of B_s and $\overline{B}_s$

As explained in Section 1.4.2, in the untagged B_s study, we take the reduced angular distribution because the two (interference) terms that have opposite signs for particle and anti-particle cancel each other, assuming equal number of B_s and $\overline{B}_s$. However, in a sample of a limited size a statistical fluctuation enhancing (anti-)particle content may be large enough leading to an incomplete cancellation and, consequently, a systematic effect.

The dispersion of the binomial distribution for N Bernoulli trials with probability of success p is $Np(1-p)$. So in our sample of about $N = 200$ candidates, of which half are expected to be B_s (and the other half – $\overline{B}_s$) the size of binomial fluctuation is $\sqrt{200 \cdot \frac{1}{2} \cdot \frac{1}{2}}$, *i.e.* about 7%. In the presence of such fluctuation in the number of B_s and $\overline{B}_s$ results in terms 4 and 6 in Equation 1.77 not canceling exactly, but leaving a remnant of size of about 7%.

However, the time dependence of the considered interference terms includes the form $\sin(\delta - \Delta M_s t)e^{-\Gamma t}$. In the fit, they get convoluted with a resolution function of width $\sigma_{ct} \approx 30\,\mu$m, and if the sinusoidal oscillations are rapid enough, additional suppression of these terms takes place. To estimate this suppression quantitatively we can make use of $\sigma_{ct} \ll c/\Gamma$ in approximating the convolution:

$$\begin{aligned} P(t) &= \int \sin(\delta - \Delta M_s t')e^{-\Gamma t'} G(t-t', \sigma_{ct})dt' \\ &\approx e^{-\Gamma t} \int \sin(\delta - \Delta M_s t') G(t-t', \sigma_{ct})dt' \\ &\sim e^{-\frac{1}{2}\Delta M_s^2 \sigma_{ct}^2} \sin(\delta - \Delta M_s t)e^{-\Gamma t}. \end{aligned} \tag{3.41}$$

For $\Delta M_s = 20\,\text{ps}^{-1}$ and $\sigma_{ct} = 0.1\,\text{ps}$, the suppression factor in front of $\sin(\delta - \Delta M_s t)e^{-\Gamma t}$ is roughly $e^{-2} \approx 0.13$. Thus the effective fraction of the signal that weights these terms is (starting from a possible 7% fraction from above) $0.07 \times 0.13 = 0.01$.

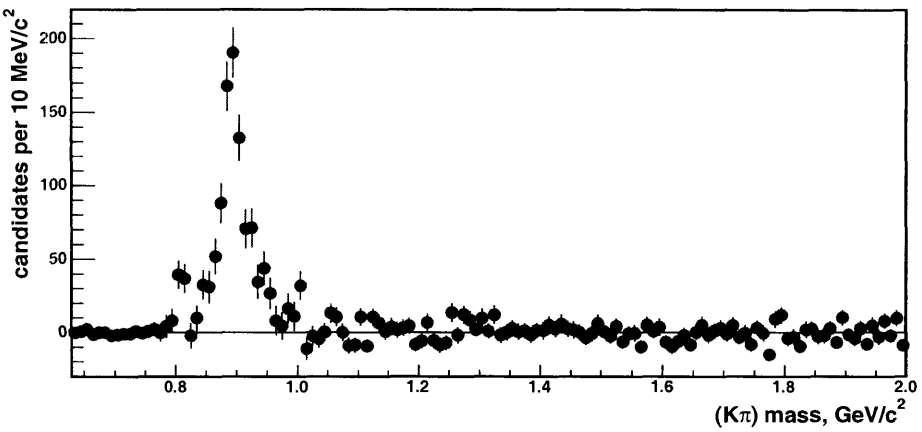


Figure 3-25: Spectrum of the invariant mass of the $K\pi$ combinations used in reconstructing B_d signal candidates. Background subtraction is performed using B_d mass side-bands.

Since this is such a small fraction, we do not assign a systematic uncertainty due to potentially unequal number of B_s and $\overline{B}_s$ in the sample.

3.4.8 Non-resonant contribution to $\mu^+\mu^- K^+\pi^-$ final state

Some of the reconstructed B_d signal may come from $\mu^+\mu^- K^+\pi^-$ final states that are not achieved via $K^{*0}(892)$ resonance. Most likely contributors are the direct decay $B_d \to J/\psi K^+\pi^-$, which has a phase-space distribution for the $K\pi$ system and B_d decays to J/ψ and a wide K^{*0} resonance centered above $1.4\,\mathrm{GeV/c^2}$ and decaying to $K^+\pi^-$. To illustrate this situation, in Figure 3-25 we plot the invariant mass of $K\pi$ combinations resulting from B_d candidates reconstructed without $|m^{K\pi} - m^{K^{*0}}_{PDG}| < 50\,\mathrm{MeV/c^2}$ requirement.

Figure 3-25 suggests no substantial contamination of this sort. The BaBar Collaboration has estimated [56] the potential contamination under the $K^{*0}(892)$ peak at $(1.2\pm0.7)\%$ by adding a term describing the angular shape of decays other than those proceeding through $K^{*0}(892) \to K^+\pi^-$ and allowing appropriately for the interference. Our naive estimate with Monte-Carlo simulation using the branching fractions from Reference [25] gives a conservative upper bound of 4%. With these numbers giving the size of the contamination, we can estimate the magnitude of the corresponding systematic effect in much the same way we did for the case of B_d cross feed into the B_s sample, *i.e.* by performing the fit on a Monte-Carlo sample with and without contaminating component. The estimates of the systematic uncertainties resulted from doing this exercise are given in Table 3.18.

3.4.9 Other sources

There are potentially other systematic effects in the reconstruction affecting the lifetime measurements due to:

Par.	$\|A_0\|^2$	$\|A_\|\|^2$	$\|A_\perp\|^2$	$\delta_\|$	$\delta_\perp$
Syst. uncert. assigned	± 0.010	± 0.001	± 0.003	± 0.07	± 0.04

Table 3.18: Systematic uncertainties assigned due to potential contribution of the non-resonant decays to the signal component in the B_d sample.

- How the primary vertex (PV) position is determined. We take the beam position at z coordinate of the J/ψ vertex, but the beam tilt ($\partial z/\partial x, \partial z/\partial y \neq 0$) is not taken into account (it is easy to see though with the analytical calculation that the maximum possible effect is below 0.1 μm).
- Magnetic field is not parallel to the beam, is not uniform, *etc.*
- SVX efficiency can be a little higher for large L_{xy} because of slightly stronger correlation (as "viewed" by the SVX) among B daughters.

For example, the generated lifetime for the results in Table 3.8 was 496.00 μm, whereas we extract 497.38 ± 1.01. In order to see if there really is a reconstruction effect behind this, we turn to our largest (80M generated) B_d Monte-Carlo sample (Sample 2 from Appendix C). In Figure 3-26 we plot the difference of the reconstructed PDL and the generated one on a candidate-by-candidate basis. We fit the resultant distribution to a triplet of Gaussians with a common mean.

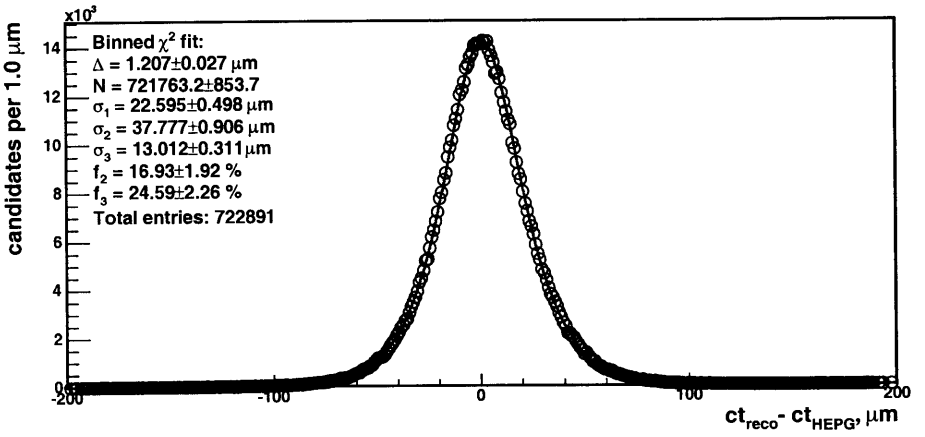


Figure 3-26: The difference between the reconstructed and the generated PDL. $B_d \rightarrow J/\psi K^{*0}$ Monte-Carlo sample.

The mean is at 1.2 μm, which we take as a systematic uncertainty due to L_{xy} and p_T measurements, and potential detector/reconstruction effect.

3.4.10 Summary

In the summary of this section we collect all estimates of the systematic uncertainties pertaining to different analyses in the respective tables and calculate the total systematic uncertainty for each measured quantity.

For the average lifetime measurements the results are given in Table 3.19. The systematic uncertainties of the results obtained in the time-dependent angular analysis of $B_d \to J/\psi K^{*0}$ and $B_s \to J/\psi\phi$ decays are summarized in Tables 3.20 and 3.21.

Systematic effect	Uncertainty on		
	$c\tau_{B_u}$, μm	$c\tau_{B_d}$, μm	$c\tau_{B_s}$, μm
SVX alignment	± 1.0		
Fit model	± 1.7		
Selection	—		
Detector bias	-1.2		
Cross feed	—	—	$-1.7\,\mu$m
Total	$^{+2.0}_{-2.4}$	$^{+2.0}_{-2.4}$	$^{+2.0}_{-2.9}$

Table 3.19: Summary of the systematic uncertainties assigned to the average lifetime measurements.

Systematic effect	Uncertainty on					
	$c\tau$, μm	$\lvert A_0\rvert^2$	$\lvert A_\parallel\rvert^2$	$\lvert A_\perp\rvert^2$	$\delta_\parallel$	$\delta_\perp$
Bkg. angular model	± 3.9	± 0.013	± 0.006	± 0.007	± 0.01	± 0.01
Eff. and acc.	—	—	—	—	—	—
K $\leftrightarrow$ π swap	—	± 0.006	± 0.004	± 0.002	± 0.04	—
Non-resonant decays	—	± 0.010	± 0.001	± 0.003	± 0.07	± 0.04
Lft. fit model	± 1.7	—	—	—	—	—
SVX alignment	± 1.0	—	—	—	—	—
Detector bias	-1.2	—	—	—	—	—
B_s cross feed	—	—	—	—	—	—
Total	$^{+4.4}_{-4.6}$	± 0.017	± 0.007	± 0.007	± 0.08	± 0.04

Table 3.20: Systematic uncertainties assigned to the quantities measured in the time-dependent angular analysis of $B_d \to J/\psi K^{*0}$.

3.5 Cross checks

The evidence of the two lifetimes in the B_s sample is a non-trivial result, especially if the central value for $\Delta\Gamma_s/\Gamma_s$ stays where currently measured while the uncertainty is reduced (by adding more statistics, combining with a measurement by the DØ Collaboration once it comes out, *etc.*). For the reason that most scrutiny is likely to be aimed at this particular result of the dissertation, it is the one most important to cross-check.

Systematic effect	Uncertainty on					
	$c\tau_L$, μm	$\Delta\Gamma/\Gamma$	$\lvert A_0\rvert^2$	$\lvert A_\parallel\rvert^2$	$\lvert A_\perp\rvert^2$	$\delta_\parallel$
Bkg. angular model	± 3.7	± 0.007	± 0.011	± 0.013	± 0.002	± 0.03
Eff. and acc.	—	—	—	—	—	—
Unequal # of B_s and $\overline{B}_s$	—	—	—	—	—	—
Lft. fit model	± 1.7	—	—	—	—	—
SVX alignment	± 1.0	—	—	—	—	—
Detector bias	-1.2	—	—	—	—	—
B_d cross feed	-5.0	± 0.008	—	± 0.003	± 0.003	—
Total	$^{+4.2}_{-6.7}$	± 0.011	± 0.011	± 0.013	± 0.004	± 0.03

Table 3.21: Systematic uncertainties assigned to the quantities measured in the time-dependent angular analysis of $B_s \to J/\psi\phi$.

3.5.1 Fitter tests

Pulls. Per standard procedure of making sure that the fitter does not introduce a bias in any of the extracted quantities of interest, we perform fits on a large number of Toy Monte-Carlo samples (also known as pseudo-experiments) and compare the output to the input by means of constructing *pull* distributions. In our case a pull for parameter P is defined as $(P_{fit} - P_{input})/\sigma_{P_{fit}}$. In the ideal case pull distribution should be a unit width Gaussian centered at zero.

In the first exercise we perform the fit on a 1000 Toy Monte-Carlo samples of 12380 events in each. The events are generated such that they mimic the properties of the B_s data sample. In this exercise we generate ten times more events per experiment than we have in the actual data sample, so this is mostly a test of the fitter as a machine to correctly extract the numbers encoded in the Toy Monte-Carlo sample. This exercise did not reveal anything we think we should worry about.

In the next exercise we fit 7585 Toy Monte-Carlo samples with 1238 events in each. This potentially tells us if there is any significant change in the fitter behavior that might be caused by the lower statistics of the actual B_s data sample (1238 events). Similar to what has been observed in the past, *e.g.* in the lifetime analysis [105], the pulls for the lowest statistics components of the sample, such as (f_{++}, λ_{++}) start showing non-Gaussian behavior. However, for the two parameters that we are most interested in, $c\tau_L$ and $\Delta\Gamma/\Gamma$, we get pull distributions quite compatible with a unit width Gaussian centered at zero, as can be seen in Figure 3-27.

All other fitters used in this dissertation are also checked in a similar fashion with no problem found.

'Fitting' for $\Delta\Gamma/\Gamma$ using the B_d sample. In order to make sure that the presence of the background with potentially complex angular distribution does not somehow make the fitter "manufacture" a non-trivial $\Delta\Gamma/\Gamma$ we consider the following exercise. We apply the B_s fitter to the B_d sample, where we expect to see no lifetime difference. The B_d sample used in the time-dependent angular analysis is about four times as big as the B_s sample, so in addition to the fit on the entire B_d sample, we split the events into four separate sub-samples and perform the test on each of

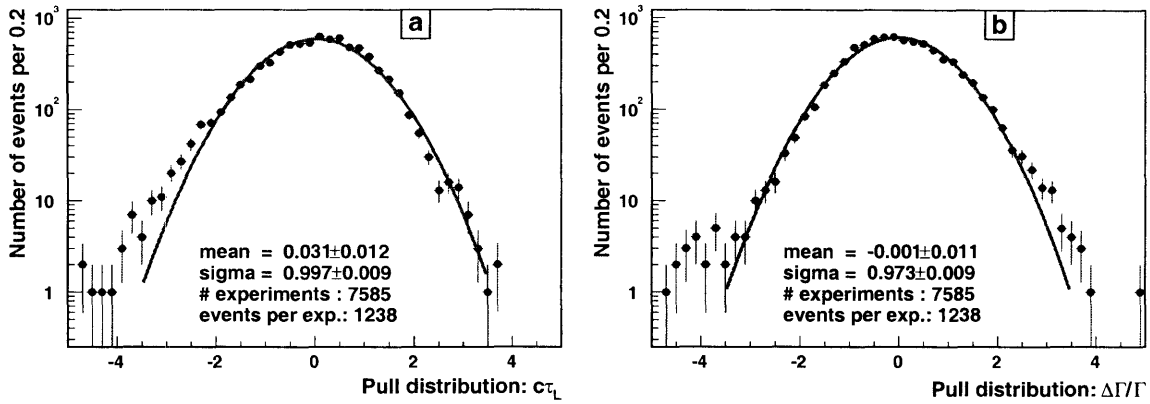


Figure 3-27: Pull distributions for the $c\tau_L$ and $\Delta\Gamma/\Gamma$ parameters obtained from 7585 Toy Monte-Carlo experiments with B_s fitter.

Sample	$\Delta\Gamma/\Gamma$, %	$c\tau_L$, μm
Full sample, one lifetime	**0.0** (fixed)	461 ± 15
Full sample	14.5 ± 12.1	444 ± 21
Sub-sample A	13.7 ± 27.9	422 ± 34
Sub-sample B	25.1 ± 22.3	437 ± 39
Sub-sample C	26.1 ± 23.0	437 ± 50
Sub-sample D	-7.6 ± 27.6	475 ± 41

Table 3.22: Results of applying B_s fit model with $\Delta\Gamma/\Gamma$ to the B_d sample.

them. Table 3.22 lists the $\Delta\Gamma/\Gamma$ results, all of which are compatible with zero. It is important to note that these results *do not* represent a measurement or limit on the size of a potential lifetime difference in the B_d system; as discussed earlier our data are insensitive to such an effect. This test is really a technical check that we obtain zero when we expect to do so.

3.5.2 CP-odd fraction versus the PDL cut value

A cut on the proper decay length is expected to enhance the longer lived CP-odd component. The larger the difference between the lifetimes of CP-odd and CP-even components the more pronounced should be the enhancement. In order to see if the data conform to this prediction we fit for the CP-odd fraction using different ct cuts. When fitting with a ct cut we exclude time-dependence from the likelihood function, *e.g.* we only use mass and angular information as in Reference [50]. The CP-odd fractions extracted with such a method from the B_s sample with ct cuts at $0\,\mu$m, $150\,\mu$m, $300\,\mu$m, and $450\,\mu$m are listed in the second column of Table 3.23. The increase of the CP-odd fraction suggests a relatively large lifetime difference which is consistent with our result. We may also predict the fraction using our measurements of $c\tau_L$ and $c\tau_H$ and assuming CP-odd fraction of 20.1% at $ct > 0$ cut. These predictions are listed in the third column of the above table. Although the uncertainties are

$ct >$	(C)P-odd fraction, %		
cut, μm	B_s, fitted	B_s, predicted	B_d, fitted
0	20.1 ± 9.0	20.1 (reference)	21.6 ± 4.4
150	24.2 ± 10.3	24.1	23.0 ± 3.6
300	29.6 ± 12.7	28.6	23.0 ± 4.0
450	38.7 ± 11.6	33.6	23.6 ± 4.9

Table 3.23: (C)P odd fraction vs. lifetime cut.

large, the agreement between the measurement and the prediction is consistent.

Finally, to get a feeling of how significant these results are we fit for the P-odd fraction in the B_d sample using the same sequence of *ct* cuts. The expectation is that it will remain stable, no matter what the *ct* cut is. Indeed, as column 4 of Table 3.23 shows, the P-odd fraction in the B_d sample *does not* change as a function of *ct* cut applied.

3.5.3 Constraining the B_s amplitudes

One might expect from the flavor SU(3) symmetry that in the fit for $\Delta\Gamma_s/\Gamma_s$ it is reasonable to fix the B_s transversity amplitudes to those measured in $B_d \to J/\psi K^{*0}$; the reason being that the latter are more precisely known or that an independent measurement of these can be used. Though the degree to which the flavor SU(3) symmetry makes the two sets of amplitudes agree is not known, we perform additional fits on the B_s sample:

- By fixing the B_s transversity amplitudes to the values found in Reference [56] we get $\Delta\Gamma_s/\Gamma_s = 0.56 \pm 0.19$.

- By fixing the B_s transversity amplitudes to the values found in Reference [56] *and* constraining $\Gamma_s = \Gamma_d$ to within 1% we get $\Delta\Gamma_s/\Gamma_s = 0.52 \pm 0.18$.

It appears that even if using as much external input as possible the evidence for a large width difference remains.

3.5.4 Single lifetime fit in the B_s sample

It has been already alleged in the summary of Section 3.3.1, that the average lifetime measured in the $B_s \to J/\psi\phi$ sample is suggestive of a large width difference between the mass eigenstates in the $B_s - \overline{B}_s$ system.

Indeed, in a semi-quantitative exercise we can take $|A_\perp|^2$ measurement for B_d from References [56, 57], and by arguments from $SU(3)$ symmetry, claim that $|A_\perp|^2$ in B_s is about the same, 0.16. This means that the number of CP-odd decays in the B_s sample considered is proportional to $0.16\tau_H$ and the number of CP-even decays is proportional to $0.84\tau_L$. Using the result of Appendix A and 408.8 μm as the average

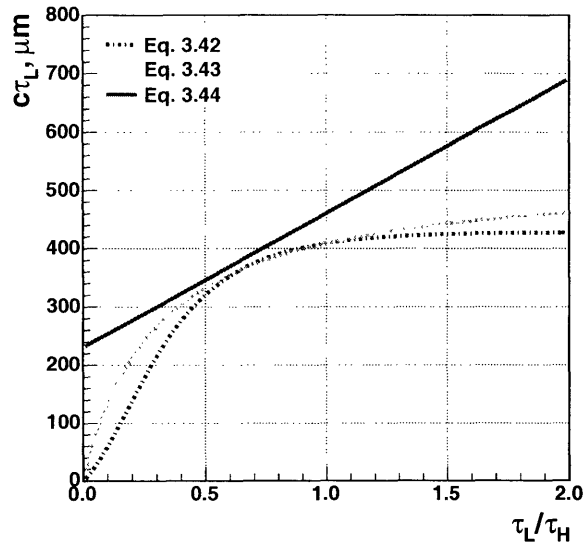


Figure 3-28: $c\tau_L$ vs. τ_L/τ_H using *only* the average lifetime measurement in the $B_s \to J/\psi\phi$ sample and an *independent* estimate of the composition of the sample in terms of CP fractions.

lifetime measurement in B_s sample (Section 3.3.1) we may assert that

$$\frac{0.16\tau_H}{0.16\tau_H + 0.84\tau_L} c\tau_H + \frac{0.84\tau_L}{0.16\tau_H + 0.84\tau_L} c\tau_L = 408.8\,\mu\text{m}. \tag{3.42}$$

Alternatively, we can take a less precisely measured time-integrated fraction of the CP-odd component from Reference [50], and declare that 23% of the decays in the sample are CP-odd, while the other 77% are CP-even to get

$$0.23\, c\tau_H + 0.77\, c\tau_L = 408.8\,\mu\text{m}. \tag{3.43}$$

At the same time, if $\Gamma_s = \Gamma_d$ holds, we have

$$\frac{2c\tau_H\tau_L}{\tau_H + \tau_L} = 460.8\,\mu\text{m}. \tag{3.44}$$

Each of the Equations 3.42 – 3.44 can be re-written as $c\tau_L$ being a function of τ_L/τ_H. Having done so we plot the corresponding functions in Figure 3-28 and interpret the point of the nearest approach of the lines corresponding to Equations 3.42 and 3.44 (or Equations 3.43 and 3.44) in terms of $c\tau_L$ and $\Delta\Gamma/\Gamma$ to get approximately 350 μm and 50% in agreement with the results of time-dependent angular analysis.

3.5.5 Estimated number of candidates at large PDL

Based on the PDL projection plot Figure 3-17 (b), it appears that there might be more events at very long PDLs than predicted by the fit. Another concern might be that $\Delta\Gamma/\Gamma$ could be artificially manufactured by the excess of background events

Component	$N_{cand.}$ beyond			
	$ct = 2000\,\mu$m		$ct = 1000\,\mu$m	
	Expected	Observed	Expected	Observed
CP-odd+CP-even	1.8+0.3	2.2	8.9+6.7	11.2
Background	0.5	1.2	2.3	3.6

Table 3.24: Estimate of the actual composition of the longer lived part of the B_s sample and the expectation based on the fit results. The observed numbers are corrected for the fact that beside 203 signal candidates there are about 160 background candidates in the signal region. This is the source of non-integer number of observed candidates.

at large PDLs. To quantitatively check these, we use the central values for the time-integrated CP fractions and other quantities reported in Tables 3.6 and 3.7 to calculate the number of events beyond a certain PDL point and then compare those to what we actually see in data. When looking at data, we define the signal region as $\pm 24\,\mathrm{MeV/c^2}$ around $5366\,\mathrm{MeV/c^2}$ in mass and everything outside as the background region.

Table 3.24 gives the comparison of the prediction based on fit results and the observation estimate based on where candidates are in the mass distribution. Given that the "Expected" numbers have several tens of percent relative uncertainty on them, we conclude that composition of the longer lived part of the sample is in agreement with the model.

3.6 Summary and discussion

Here we quote the final results of all of the analyses described in this dissertation. Following the standard convention, the first uncertainty quoted is statistical and the second one is systematic. The results are followed by the discussion, comparison to other measurements, as well as a review of the most important implications.

3.6.1 Final results

For the measurement of the average lifetimes we obtain:

$$\begin{aligned} \tau_{B_u} &= (1.659 \pm 0.033\ ^{+0.007}_{-0.008})\,\mathrm{ps}, \\ \tau_{B_d} &= (1.549 \pm 0.051\ ^{+0.007}_{-0.008})\,\mathrm{ps}, \\ \tau_{B_s} &= (1.363 \pm 0.100\ ^{+0.007}_{-0.010})\,\mathrm{ps}. \end{aligned}$$

These lifetimes correspond to the following lifetime ratios:

$$\begin{aligned} \tau_{B_u}/\tau_{B_d} &= 1.071 \pm 0.042, \\ \tau_{B_s}/\tau_{B_d} &= 0.880 \pm 0.071. \end{aligned}$$

In calculating the uncertainty on the ratios we account for the fact that most systematic uncertainties are correlated and therefore nearly cancel in the ratios. They are also very small compared to the statistical ones, so it makes sense to quote a single uncertainty without splitting it into statistical and systematic parts.

In the time-dependent angular analysis of the decays of neutral B mesons to the VV intermediate states we obtain:

- For B_d:

$$\begin{aligned} |A_0|^2 &= 0.562 \pm 0.025 \pm 0.017, \\ |A_{\|}|^2 &= 0.223 \pm 0.032 \pm 0.007, \\ |A_{\perp}|^2 &= 0.215 \pm 0.032 \pm 0.007, \\ \delta_{\|} &= 2.86 \pm 0.22 \pm 0.08, \\ \delta_{\perp} &= 0.15 \pm 0.15 \pm 0.04. \end{aligned}$$

- For B_s without the $\Gamma_s = \Gamma_d$ constraint:

$$\begin{aligned} |A_0|^2 &= 0.615 \pm 0.064 \pm 0.011, \\ |A_{\|}|^2 &= 0.260 \pm 0.086 \pm 0.013, \\ |A_{\perp}|^2 &= 0.125 \pm 0.066 \pm 0.004, \\ \delta_{\|} &= 1.93 \pm 0.36 \pm 0.03, \\ \tau_L &= (1.05 \,^{+0.16}_{-0.13} \pm 0.02)\,\text{ps}, \\ \tau_H &= (2.07 \,^{+0.58}_{-0.46} \pm 0.03)\,\text{ps}, \\ \Delta\Gamma/\Gamma &= 0.65 \,^{+0.25}_{-0.33} \pm 0.01, \\ \Delta\Gamma &= (0.47 \,^{+0.19}_{-0.24} \pm 0.01)\,\text{ps}^{-1}. \end{aligned}$$

- For B_s with the $\Gamma_s = \Gamma_d$ constraint:

$$\begin{aligned} |A_0|^2 &= 0.614 \pm 0.064 \pm 0.011, \\ |A_{\|}|^2 &= 0.291 \pm 0.080 \pm 0.013, \\ |A_{\perp}|^2 &= 0.095 \pm 0.052 \pm 0.004, \\ \delta_{\|} &= 1.90 \pm 0.32 \pm 0.03, \\ \tau_L &= (1.13 \,^{+0.13}_{-0.09} \pm 0.02)\,\text{ps}, \\ \tau_H &= (2.38 \,^{+0.56}_{-0.43} \pm 0.03)\,\text{ps}, \\ \Delta\Gamma/\Gamma &= 0.71 \,^{+0.24}_{-0.28} \pm 0.01, \\ \Delta\Gamma &= (0.46 \pm 0.18 \pm 0.01)\,\text{ps}^{-1}. \end{aligned}$$

3.6.2 Comparisons to other results

Figures 3-29–3-31 allow one to compare our results for the average lifetimes to the PDG averages [25] and single best measurements. Our measurement of the average B_s lifetime is the single best measurement in terms of the overall uncertainty. Given

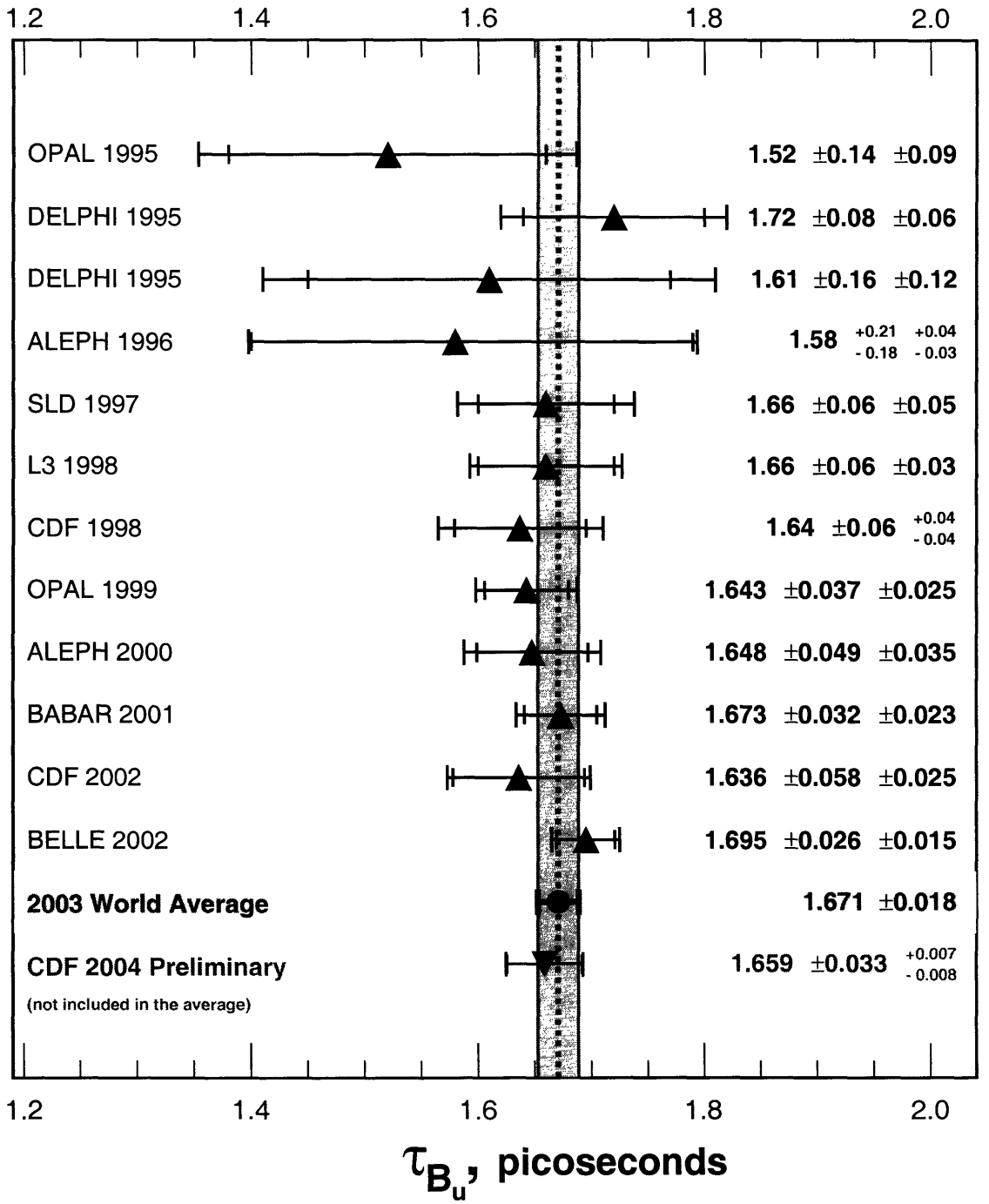


Figure 3-29: 2003 World Average of the B_u lifetime and the measurements contributing to it. The result of the B_u lifetime measurement described in this dissertation is labeled "CDF 2004 Preliminary" and is shown at the bottom of the plot.

the rate at which statistics is accumulated, B_u lifetime measurement is approaching such a honorable status.

A comparison of the angular decay amplitudes to other measurements is shown

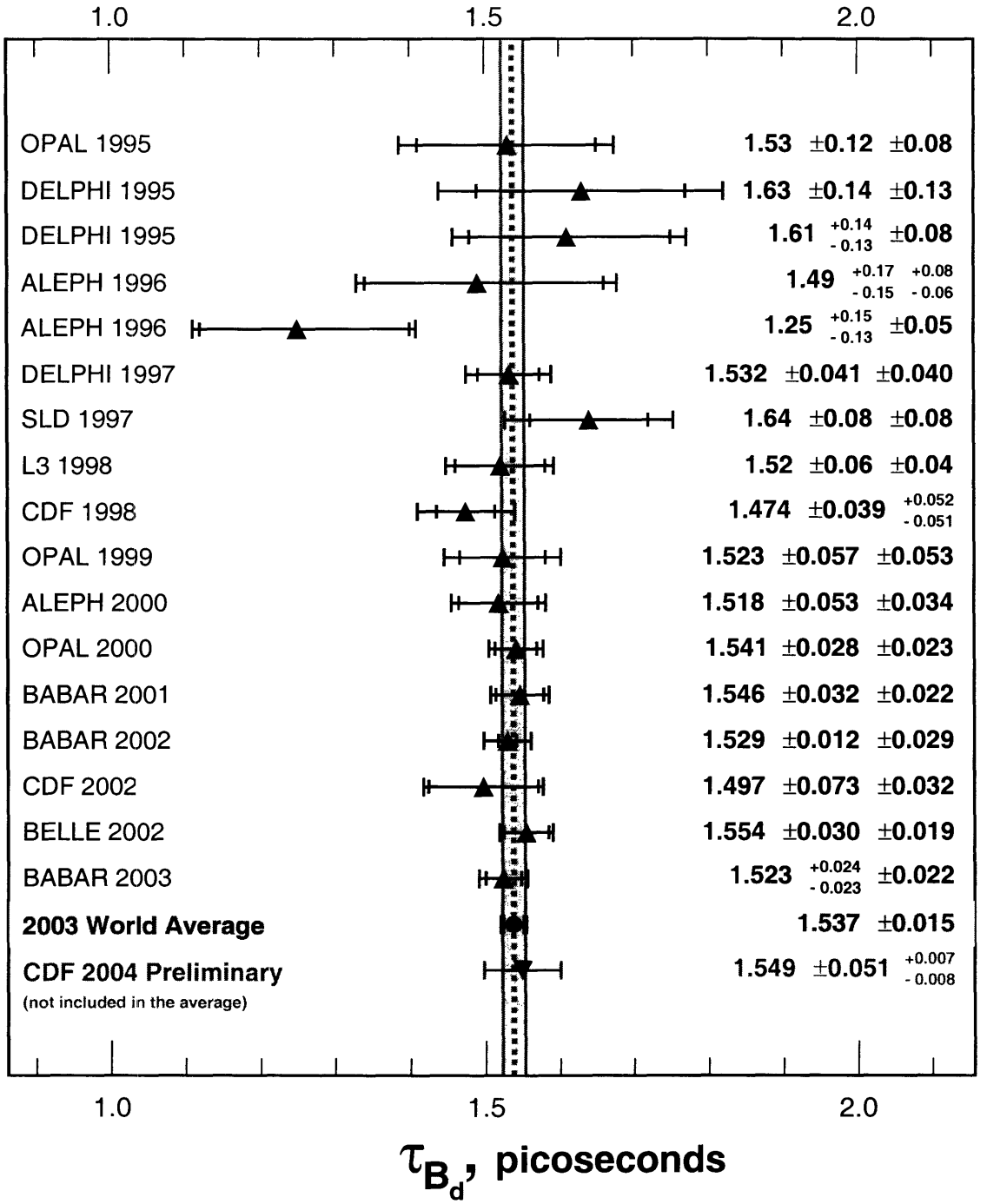


Figure 3-30: 2003 World Average of the B_d lifetime and the measurements contributing to it. The result of the B_d lifetime measurement described in this dissertation is labeled "CDF 2004 Preliminary" and is shown at the bottom of the plot.

in Figure 3-32. Table 1.5 can be used for numerical comparison. We have competitive and comparable results for the B_d amplitudes and phases to those published by BaBar [56] and Belle [57]. The amplitudes measured in the $B_s \to J/\psi\phi$ decay

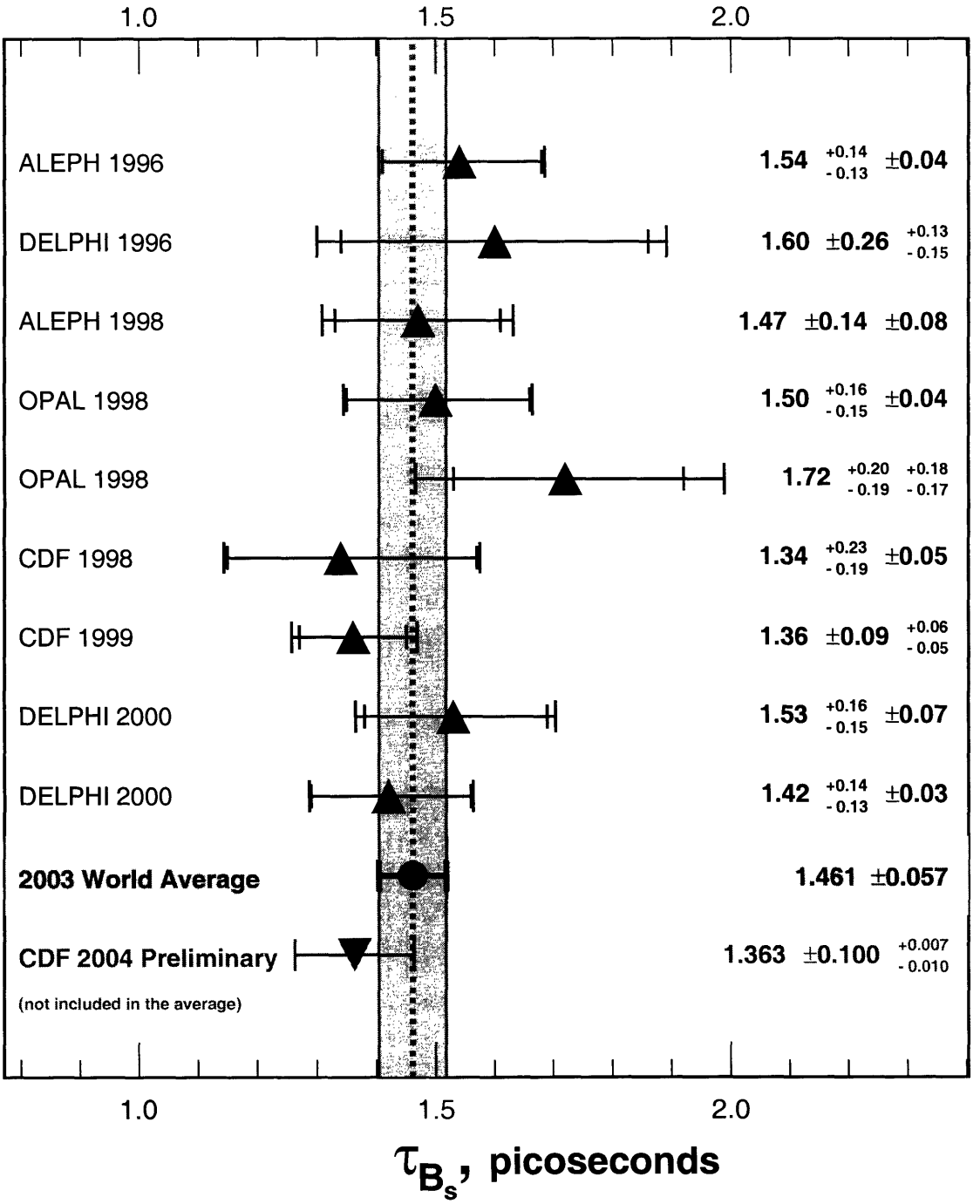


Figure 3-31: 2003 World Average of the B_s lifetime and the measurements contributing to it. The result of the average B_s lifetime measurement described in this dissertation is labeled "CDF 2004 Preliminary" and is shown at the bottom of the plot. In the presence of a sizable $\Delta\Gamma_s$ the direct comparison of the mean values for the B_s lifetime measurements is not meaningful, for those measurements are performed on different mixtures of heavy and light mass eigenstates.

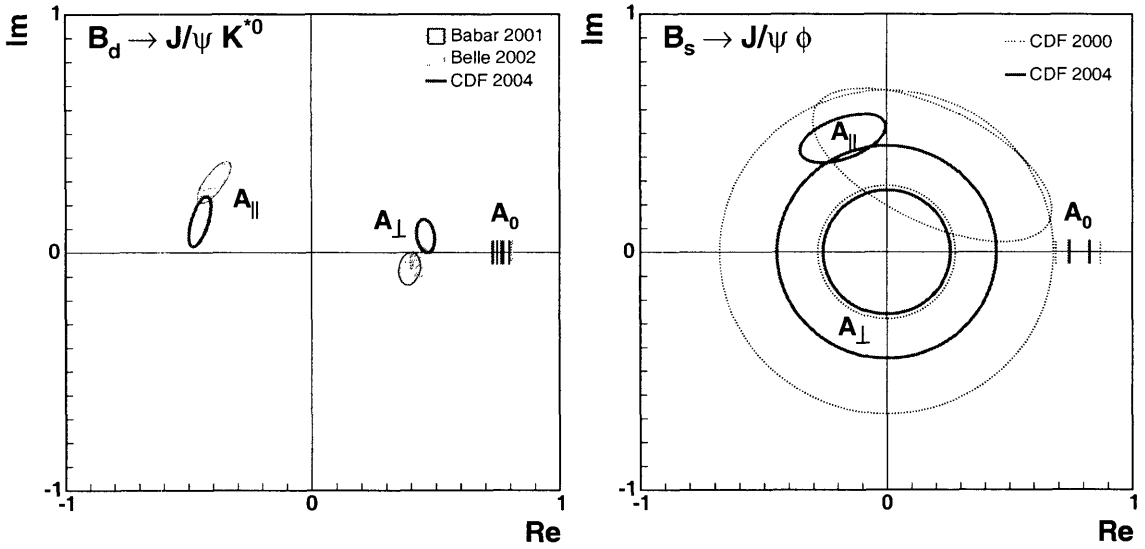


Figure 3-32: Angular decay amplitudes measured in (a) $B_d \rightarrow J/\psi K^{*0}$, (b) $B_s \rightarrow J/\psi\phi$. The uncertainty drawn reflects the statistical and systematic uncertainties added in quadrature. Correlations between the phase and the magnitude of the amplitude in $A_{\parallel,\perp}$ are neglected. One standard deviation contours are plotted.

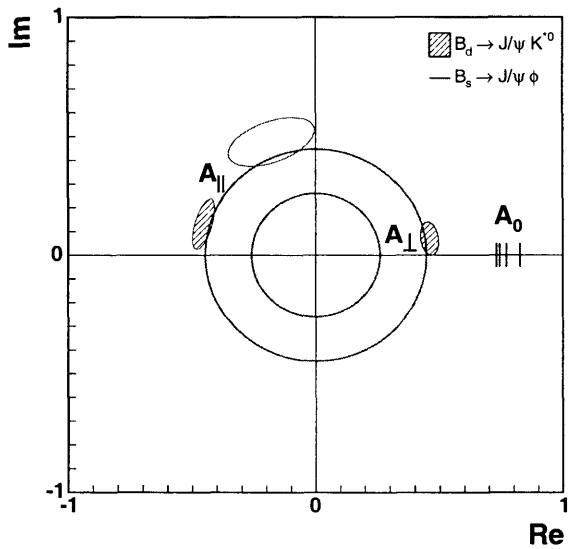


Figure 3-33: Angular decay amplitudes for B_d and B_s overlaid on the same plot.

agree with the Run I measurement [55], but are substantially more precise. The gain in precision comes not only from the increased statistics, but also from using proper decay length information which allows one to separate signal from background much more effectively than the mass information alone.

The amplitudes and, to a lesser degree, phase $\delta_\parallel$ are in agreement, between B_d and B_s, as illustrated by Figure 3-33.

3.6.3 What are the betting odds?

An interesting question to ask is how likely it would be for an ensemble of experiments with statistical sensitivity similar to that of our analysis to observe a value for $\Delta\Gamma_s/\Gamma_s$ as large as we observed if Nature provides us with no lifetime difference, $\Delta\Gamma_s/\Gamma_s = 0$, or the central value of the Standard Model prediction of $\Delta\Gamma_s/\Gamma_s = 0.12$ [30]. A standard method for answering this question is to use Monte-Carlo methods to conduct enough pseudo-experiments to obtain sufficient statistics such that the distribution of extracted $\Delta\Gamma_s/\Gamma_s$ is determined for the input value of $\Delta\Gamma_s/\Gamma_s$ we are interested in. Once this distribution is known, it is a matter of counting to determine how much of an outlier a particular observed result is. This is a non-trivial exercise, first because of the complexity of this analysis, and second because not all pseudo-experiments yield convergent fits[4].

The first issue can be to a large degree addressed by recognizing that this sort of question could be equally valid if asked of an idealized experiment, one without background or sculpting. Asking and then answering the question in this way makes things much more tractable. Nevertheless, the fit is still sufficiently complicated and, for the (still) low sample size we have (205 signal events) and correspondingly large fluctuations, convergence is not guaranteed. We have found that the rate of convergence for idealized pseudo-experiments improves greatly once the sample size is doubled and is no longer an issue when the sample size is quadrupled.

Keeping this in mind, we provide the results obtained in eight sets of 10K pseudo-experiments. In four sets we generate and fit samples of 205 pure signal events; in the other four sets we generate 1238 events per pseudo-experiment with each event having a 16.5% chance to be signal and a 83.5% chance to be background. In all pseudo-experiments signal is generated with $|A_\perp|^2 = 0.22$. For this exercise we assume the value for $|A_\perp|^2$ observed in the B_d system where it is more precisely measured. Also, given the input value of $\Delta\Gamma_s/\Gamma_s$, the lifetime of the light mass eigenstate, $c\tau_L$, is generated using $\Gamma_s = \Gamma_d = c/460.8\,\mu$m. The betting odds and equivalent Gaussian significances are compiled in Table 3.25. They are based on the fits that converge.

For comparison, the $\sqrt{-2\log(\mathcal{L}/\mathcal{L}_{max})}$ values obtained from the likelihood scans shown in Figure 3-19 are:

- 1.60 at $\Delta\Gamma_s/\Gamma_s = 0$ for the unconstrained fit,
- 2.30 at $\Delta\Gamma_s/\Gamma_s = 0$ for the constrained fit,
- 1.41 at $\Delta\Gamma_s/\Gamma_s = 0.12$ for the unconstrained fit,
- 1.99 at $\Delta\Gamma_s/\Gamma_s = 0.12$ for the constrained fit.

These values are lower by about 0.4(1.0) for the (un)constrained fit than the corresponding values obtained with pseudo-experiments. The reason is that these last

[4]We would like to emphasize however, that the fit to the actual data converges starting from anywhere in a rather broad range of initial values. In particular, the constrained fit converges to the quoted central value with a starting value of $\Delta\Gamma_s/\Gamma_s = 0$.

	Unconstrained fit				Constrained fit			
	Sig. only		Sig.+Bkg.		Sig. only		Sig.+Bkg.	
Input $\Delta\Gamma_s/\Gamma_s$	0.0	0.12	0.0	0.12	0.0	0.12	0.0	0.12
# of convergent fits	8179	8488	6709	7003	9333	9371	8107	8072
# of $\Delta\Gamma_s/\Gamma_s > 0.65$	26	101	20	94	—	—	—	—
# of $\Delta\Gamma_s/\Gamma_s > 0.71$	—	—	—	—	13	46	27	54
Betting odds, 1 in	315	84	335	75	718	204	300	149
Equivalent Gaussian significance, std. deviation	2.73	2.26	2.75	2.21	2.99	2.58	2.72	2.48

Table 3.25: Compilation of the estimates of how significantly away from $\Delta\Gamma_s/\Gamma_s = 0.0$ or $\Delta\Gamma_s/\Gamma_s = 0.12$ the results of the unconstrained and constrained fits are.

values give an estimate of the answer to a subtly different question, namely, how likely it is that the true $\Delta\Gamma_s/\Gamma_s$ value, being what we have measured it to be, fluctuates down to 0 or 0.12. Given that the likelihood scans for $\Delta\Gamma_s/\Gamma_s$ are not parabolic and much less steeply rising toward lower $\Delta\Gamma_s/\Gamma_s$, such pattern of values is expected.

It is also important to stress that it is the information provided by the likelihood scans that ought to be used in any global combination of the results of this dissertation with related previous or future measurements.

3.6.4 Discussion

A common feature of all of the results compiled in Section 3.6.1 is that their uncertainties are dominated by statistics. Recently, when the Tevatron operated stably, the weekly integrated luminosity was $7 - 10\,\mathrm{pb}^{-1}$. Given that, it is reasonable to expect the improvement in precision of each of the measurements by a factor of two on a time scale of a year, assuming that the CDF and DØ results are combined. At that point, if the central value for $\Delta\Gamma_s/\Gamma_s$ stays where it is measured in the current analysis, it will be in disagreement with the Standard Model based prediction at the level of more than 3 standard deviations. It is ever more interesting in light of the fact that non-Standard Model contributions could only make $\Delta\Gamma_s/\Gamma_s$ smaller.

The combined world knowledge of $\Delta\Gamma_s/\Gamma_s$ existing prior to our current result, can be expressed as two limits [112]:

- $\Delta\Gamma_s/\Gamma_s < 0.54$ (95% C.L.) or $\Delta\Gamma_s/\Gamma_s = 0.16^{+0.15}_{-0.16}$ without the $\Gamma_s = \Gamma_d$ constraint,
- $\Delta\Gamma_s/\Gamma_s < 0.29$ (95% C.L.) or $\Delta\Gamma_s/\Gamma_s = 0.07^{+0.09}_{-0.07}$ with the $\Gamma_s = \Gamma_d$ constraint imposed.

The latter, which seems to be in disagreement with the measurement of this dissertation, is greatly influenced by the B_s lifetime measured in semileptonic decays, $\tau_{B_s}^{semil.}$. However, the existing measurement of $\tau_{B_s}^{semil.}$ should be treated cautiously, at least if $\Gamma_s = \Gamma_d$ is invoked. Indeed, as discussed in Section 1.5.3, $\Gamma_s = \Gamma_d$ implies

$\tau_{B_s}^{semil.} \geqslant \tau_{B_d}$, which is *not* the case for the central values of the measurements of these quantities.

The two measurements of $\Delta\Gamma_s/\Gamma_s$ by ALEPH [60], also discussed in Section 1.5.3, do not rely on the B_s lifetime extracted from semileptonic decays. They give values of $\Delta\Gamma_s/\Gamma_s = 0.26^{+0.30}_{-0.15}$ and $\Delta\Gamma_s/\Gamma_s = 0.45^{+0.80}_{-0.49}$ [60], which are compatible with our result.

3.6.5 Implications

Evaluating all terms in Equations 1.62 and 1.63 and using Equations 1.67 and 1.68 allows one to estimate the ratio $\frac{\Delta\Gamma_s}{\Delta M_s}$ within the Standard Model [30]:

$$\frac{\Delta\Gamma_s}{\Delta M_s} = (3.7\ ^{+0.8}_{-1.5}) \times 10^{-3}.$$

To convert our measured $\Delta\Gamma_s$ into a ΔM_s value, we use the result of the constrained fit. We obtain

$$\Delta M_s = (125^{+69}_{-55})\,\mathrm{ps}^{-1},$$

which can be compared to the 95% C.L. lower limit of $\Delta M_s > 14.1\,\mathrm{ps}^{-1}$ [25].

It is also interesting to revisit the B_s lifetime measured in semileptonic decays discussed in Section 1.5.3. A simple averaging of the PDG'03 results [25] yields $c\tau_{B_s}^{semil.} = (430.8 \pm 19.8)\,\mu$m. Using Equation 1.89, we can include this information in the fit for $\Delta\Gamma_s/\Gamma_s$ in a form of a constraint, much like we did in the nominal constrained fit. Having done so, we perform the fit in terms of $c\tau_{B_s}$ and $\Delta\Gamma_s/\Gamma_s$ (see Appendix N for the transformations involved) to find $c\tau_{B_s} \equiv c/\Gamma_s = (401 \pm 29)$ ps and $\Delta\Gamma_s/\Gamma_s = 0.40 \pm 0.19$. In this way we see that the observed lifetime in B_s semileptonic decays and the evidence for non-zero width difference in the B_s system can be met if Γ_s is not equal to Γ_d, but instead is about 15% larger than Γ_d.

3.6.6 Conclusion

We see evidence for two lifetime components in the B_s system. The cross checks available, including those using the B_d system, give us confidence in this intriguing result. Our measurement of $\Delta\Gamma_s/\Gamma_s$ is not in a good agreement with the hypothesis that $\Delta\Gamma_s/\Gamma_s = 0$. We estimate that the chance of $\Delta\Gamma_s/\Gamma_s = 0$ fluctuating up to the result of the unconstrained fit is excluded with the equivalent Gaussian significance of 2.75 standard deviations. Although the value observed for $\Delta\Gamma_s/\Gamma_s$ is strikingly larger than the Standard Model prediction, $\Delta\Gamma_s/\Gamma_s = 0.12 \pm 0.06$, the statistical power of the measurement is not sufficient to claim disagreement with the Standard Model. We estimate that the result of the unconstrained fit and the central value of the Standard Model prediction are consistent within 2.21 standard deviations.

The importance of the time-dependent angular analysis described in this dissertation is not limited to just its result. It is the first time such an analysis is applied in studying B mesons, thus a large number of new challenges are addressed. This analysis also provides the first data-based benchmark of $\Delta\Gamma_s/\Gamma_s$ sensitivity, which gives

the claim that a very interesting and important B_s mixing measurement can soon be achieved at the Tevatron a much firmer ground. Last, but not least, we expect the results of this analysis to stimulate further efforts on the theoretical front.

Appendix A

Single lifetime likelihood fit

Here we prove a simple, but useful mathematical theorem. The theorem's statement is: if the lifetime is defined as parameter τ in the PDF $f(t) = \frac{1}{\tau}e^{-\frac{t}{\tau}}$, then in the maximum likelihood fit for a single lifetime for *any* sample of N events the fitted τ is simply the mean lifetime of all events: $\tau = \frac{1}{N}\sum_{i=1}^{N} t_i$.

The proof is fairly straightforward – we construct the likelihood and maximize the logarithm of it analytically:

$$\mathcal{L} = \prod_{i=1}^{N} \frac{1}{\tau} e^{-\frac{t_i}{\tau}} \tag{A.1}$$

$$\log \mathcal{L} = -N \log \tau - \frac{1}{\tau} \sum_{i=1}^{N} t_i \tag{A.2}$$

$$0 = \frac{d \log \mathcal{L}}{d\tau} = -\frac{N}{\tau} + \frac{1}{\tau^2} \sum_{i=1}^{N} t_i \tag{A.3}$$

The latter of the above equations implies that $\tau = \frac{1}{N}\sum_{i=1}^{N} t_i$ maximizes the (logarithm of the) likelihood. One can calculate the second derivative of $\log \mathcal{L}$ and plug in Equation A.3 for τ to see that the second derivative is negative and the found extremum is indeed the maximum.

Appendix B

Good run definition

To determine a list of good runs for our analyses as well as to calculate the total integrated luminosity corresponding to the data samples considered in this dissertation we use the SQL query (Listing B.1) to the database in which run information is stored. More on how this works can be found in Reference [91].

```
SELECT RUNNUMBER, sum(LUM_INTEGRAL_OFFLINE)
FROM RUNCONFIGURATIONS, FILECATALOG.CDF2_RUNSECTIONS
WHERE
RUNCONFIGURATIONS.RUNNUMBER = FILECATALOG.CDF2_RUNSECTIONS.RUN_NUMBER
AND
RUNCONFIGURATIONS.RUNCONTROL_STATUS = 1
AND
RUNCONFIGURATIONS.SHIFTCREW_STATUS = 1
AND
(RUNCONFIGURATIONS.OFFLINE_STATUS = 1
OR
RUNCONFIGURATIONS.RUNNUMBER > 175292)
AND
RUNCONFIGURATIONS.CLC_STATUS = 1
AND
RUNCONFIGURATIONS.L1T_STATUS = 1
AND
RUNCONFIGURATIONS.L2T_STATUS = 1
AND
RUNCONFIGURATIONS.L3T_STATUS = 1
AND
RUNCONFIGURATIONS.SVX_STATUS = 1
AND
RUNCONFIGURATIONS.RUNNUMBER > 138424
AND
RUNCONFIGURATIONS.RUNNUMBER < 179057
GROUP BY RUNNUMBER
/
QUIT
```

Listing B.1: SQL query (`goodrun.sql`) to the run database used in determining whether a run should be included in the analysis and what the corresponding integrated luminosity is.

Appendix C

Realistic Monte-Carlo simulation

Realistic Monte-Carlo simulation, which we most often refer to as "Monte-Carlo simulation", is an extremely important tool that finds ample use in HEP analyses. In this dissertation it is used in accounting for the sculpting of the angular decay distributions and in evaluating a number of systematic uncertainties. Below we provide a brief review of what Monte-Carlo simulation is, mention a few samples and their specific uses in this dissertation and then demonstrate the quality of the Monte-Carlo simulation we rely upon by comparing some of the generated samples with data.

A usual, at least for B physics, sequence in preparing a Monte-Carlo sample is:

1. Generating particle(s) of interest according to a known, hypothesized or a specially prepared production spectrum. Depending on the application this is done by such well-known programs, also called the *generators*, as `PYTHIA` and `HERWIG`, or less general but more suitable for a specific purpose generators, such as `Bgenerator` [107], used in the Monte-Carlo simulation for this dissertation.

2. Generating particular decay chain(s) for particle(s) produced in step 1, including appropriate kinematics for the decay products. This step is sometimes done by the same programs employed in step 1, but there are cases when one needs more flexibility than those rather general programs provide. In the present analysis we often need to generate angular decay distributions corresponding to one or another set of amplitudes, or generate a phase-space (flat) angular distributions. Almost the only program that has required functionality built in is `EvtGen` [108], originally developed at BaBar and later adopted by CDF.

3. Simulating the detector response to the events generated in step 2. This is almost universally done using the `GEANT` package [109], with some patches and modifications usually directed at making the simulation work faster. `GEANT` provides a framework for describing a rather broad range of detectors. Once the detector is built in the language of `GEANT`, almost any kind of particle can be tracked through it with all appropriate physics processes taking place so as to mimic the physical detector response. While tracking the particles, the physical detector response gets digitized (still using `GEANT`) and afterward it gets passed

to the next element of the Monte-Carlo sequence usually in a form of raw data banks.

4. Trigger simulation. At this stage raw data is fed to the algorithm that implements the actual trigger logic to decide if the event should be accepted.

5. Production. The events passing the trigger simulation go through production stage, in which the collections of physics objects (tracks, jets, muons, *etc.*) are created from the raw data as discussed in Section 2.4.

Most commonly the Monte-Carlo simulation is used to generate a single decay (chain) at a time, as this is the simplest and the least error prone way of studying the phenomena of interest. Sometimes however, generic Monte-Carlo samples including several decays or a whole class of decays are needed. One sample of the latter kind is the $b \to J/\psi(\mu^+\mu^-)X$ sample, which attempts to emulate a proper mixture of *all* B hadron decays with the $J/\psi \to \mu^+\mu^-$ decay in the chain. This sample is useful in studying reflections described in Section 3.1.5. Other major samples and their uses are listed below:

Sample 1 $B_u \to J/\psi(\mu^+\mu^-)K^+$; 10M generated, 184K reconstructed events; decay model. Purpose: cut optimization for the average lifetime analysis, comparing the Monte-Carlo simulation with data, evaluation of the systematic uncertainties.

Sample 2 $B_d \to J/\psi(\mu^+\mu^-)K^{*0}(K^+\pi^-)$; 80M generated, 725K reconstructed events; phase-space generation, *i.e.* flat in the angular space. Purpose: to study the sculpting of the angular decay distributions and determining the partial normalization constants (Section 3.2.4), also for demonstrating the agreement of the Monte-Carlo simulation with data and evaluating the systematic uncertainties.

Sample 3 $B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$; 40M generated, 516K reconstructed events; phase-space generation, *i.e.* flat in the angular space. Purpose: to study the sculpting of the angular decay distributions and determining the partial normalization constants (Section 3.2.4), evaluation of the systematic uncertainties.

Sample 4 $B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$; 7.65M generated, 101K reconstructed events; proper angular correlations, but only the CP-even amplitudes with appropriate proportion of A_0 and $A_\parallel$; B_s lifetime is set to 330 μm. Purpose: cut optimization for the $\Delta\Gamma_s/\Gamma_s$ measurement, evaluation of the systematic uncertainties.

Sample 5 $B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$; 2.35M generated, 31K reconstructed events; proper angular correlations, but the CP-odd amplitudes only; B_s lifetime is set to 550 μm. Purpose: cut optimization for the $\Delta\Gamma_s/\Gamma_s$ measurement, evaluation of the systematic uncertainties.

Sample 6 $B_d \to J/\psi(\mu^+\mu^-)K^{*0}(K^+\pi^-)$; 15M generated, 134K reconstructed events; proper angular correlations. Purpose: studying of the B_d signal reconstructed with K $\leftrightarrow$ π mass mis-assignment, evaluation of the systematic uncertainties.

As part of keeping the documentation of the analysis we collect the `EvtGen` steering files used in creating these samples at the end of this appendix.

When speaking of the size of a sample it is often useful to know how many events have been generated and how many survived the the detector simulation, trigger requirements and eventually the analysis cuts. In this dissertation when identifying a Monte-Carlo sample we quote both the number of generated events and the number of reconstructed events that have satisfied the nominal selection procedure for a given signal.

A crucial characteristic of a Monte-Carlo is how well it describes the data. A great deal of effort has been invested in the development of the Monte-Carlo framework at CDF [110], in particular in tuning the material description and achieving the resolutions close to those of the real detector. Another aspect that needs to be taken into account is that over the course of several years of data taking the real detector:

a) undergoes modifications,

b) is run under different conditions which may change as often as from run to run.

The actual datasets can therefore be rather inhomogeneous. In order to reflect that, the Monte-Carlo samples used in this dissertation are generated for a set of representative runs weighted by their luminosities.

The ultimate judgment of the quality and trustworthiness of the Monte-Carlo comes from comparison with data. We use two data samples and the corresponding two Monte-Carlo samples to make the evaluation. B_u data sample is used (in conjunction with Sample 1 from the list above) for the reason that it is the largest B meson sample and thus provides for the most accurate test. B_d data sample (in conjunction with Sample 2 from the above list) is used as the largest available to evaluate an agreement for decays with the four-track topology. Figure C-1(C-2) shows 12 distributions for the side-band subtracted $B_u(B_d)$ signal from data overlaid with their counterparts from the Monte-Carlo simulation of the same signal. Each plot shows the result of the χ^2 compatibility test. Judging from the universally good match over all of the relevant variables we conclude that the Monte-Carlo simulation describes the data at a quite adequate level.

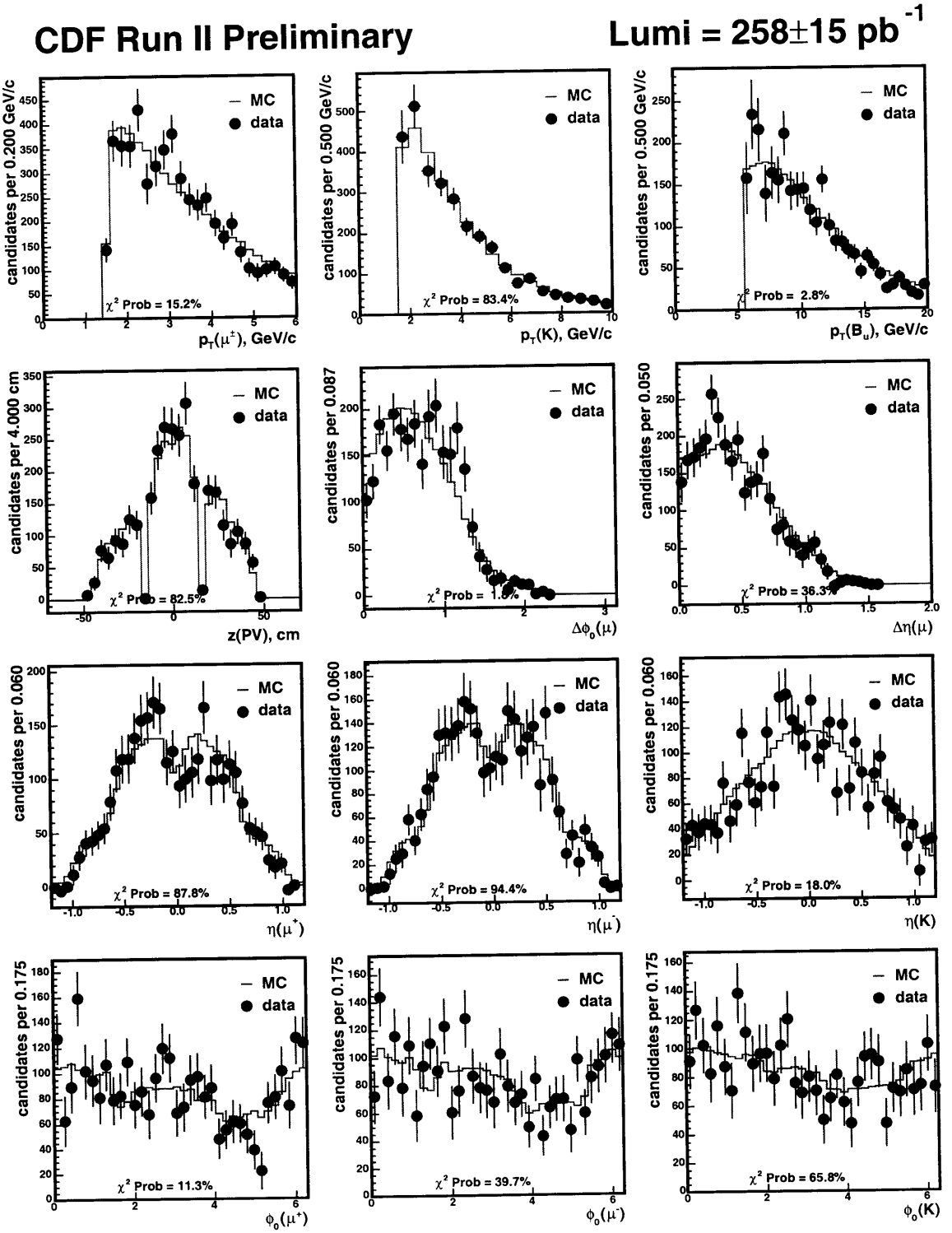


Figure C-1: B_u side-band subtracted signal from data vs. Monte-Carlo.

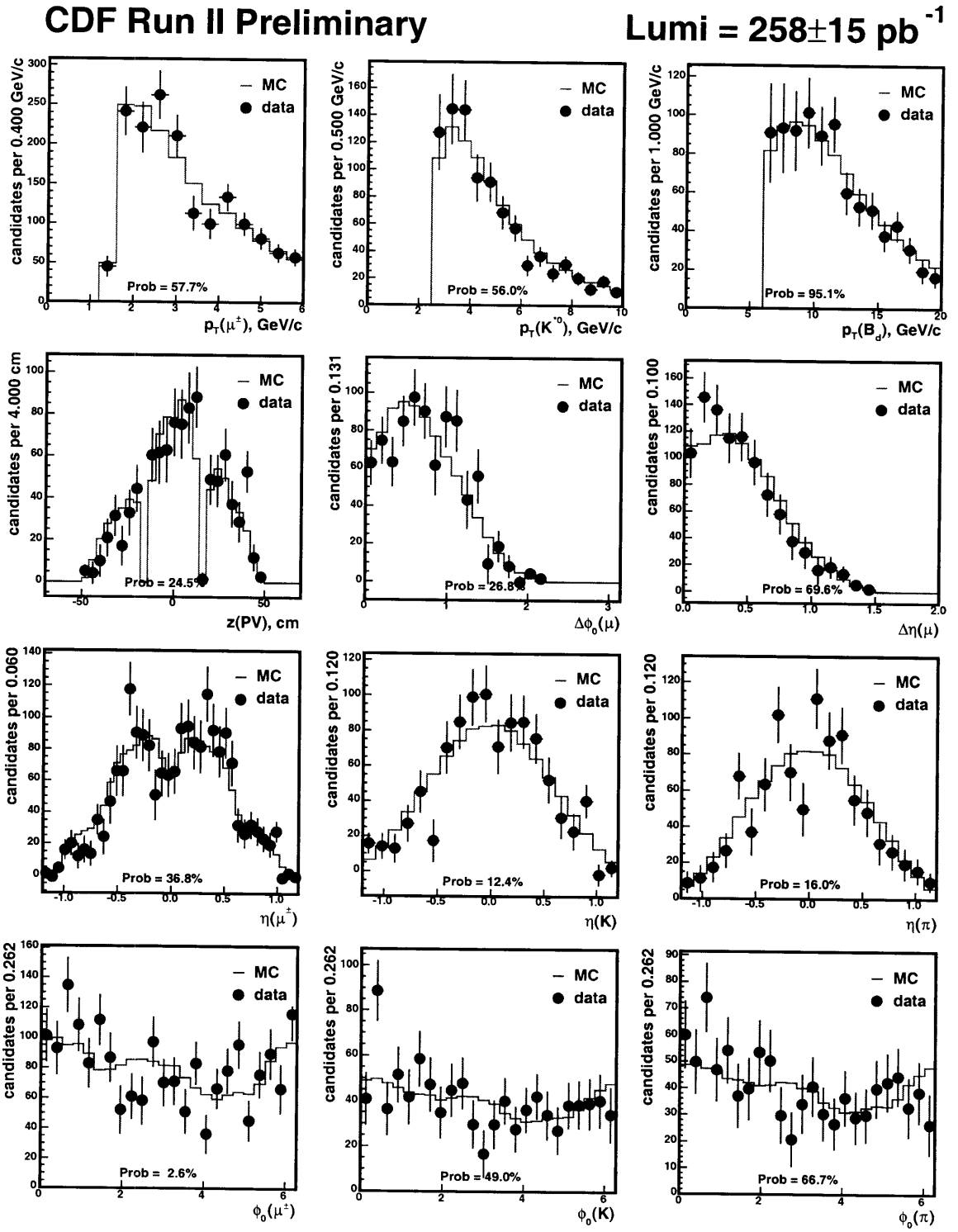


Figure C-2: B_d side-band subtracted signal from data vs. Monte-Carlo.

```
Alias MYJ/psi J/psi
Decay B+
1.000 MYJ/psi K+     SVS;
Enddecay
Decay B-
1.000 MYJ/psi K-     SVS;
Enddecay
Decay MYJ/psi
1.000 mu+ mu- PHOTOS VLL;
Enddecay
End
```

Listing C.1: $B_u \to J/\psi K^+$ sample for cut optimization for the average lifetime analysis and benchmarking the Monte-Carlo simulation.

```
Alias MYJ/psi J/psi
Alias MYK*0 K*0
Alias MYanti-K*0 anti-K*0
ChargeConj MYK*0 MYanti-K*0
Decay anti-B0
1.000 MYJ/psi MYanti-K*0 PHSP;
Enddecay
Decay B0
1.000 MYJ/psi MYK*0      PHSP;
Enddecay
Decay MYJ/psi
1.000 mu- mu+ PHOTOS     VLL;
Enddecay
Decay MYanti-K*0
1.000 pi+ K-             VSS;
Enddecay
Decay MYK*0
1.000 pi- K+             VSS;
Enddecay
End
```

Listing C.2: $B_d \to J/\psi K^{*0}$ sample for the study of the sculpting of the angular decay distributions.

```
Alias MYJ/psi J/psi
Alias MYphi   phi
Decay B_s0
1.000 MYJ/psi MYphi  PHSP;
Enddecay
Decay anti-B_s0
1.000 MYJ/psi MYphi  PHSP;
Enddecay
Decay MYJ/psi
1.000 mu+ mu- PHOTOS VLL;
Enddecay
Decay MYphi
1.000 K+ K-          VSS;
Enddecay
End
```

Listing C.3: $B_s \to J/\psi\phi$ sample for the study of the sculpting of the angular decay distributions.

```
Alias MYJ/psi J/psi
Alias MYphi  phi
Decay B_s0
1.000 MYJ/psi MYphi  SVV_HELAMP 0.363803 2.0 0.857493 0.0 0.363803 2.0;
Enddecay
Decay anti-B_s0
1.000 MYJ/psi MYphi  SVV_HELAMP 0.363803 2.0 0.857493 0.0 0.363803 2.0;
Enddecay
Decay MYJ/psi
1.000 mu+ mu- PHOTOS VLL;
Enddecay
Decay MYphi
1.000 K+ K-          VSS;
Enddecay
End
```

Listing C.4: CP-even component for p_T cut optimization for $\Delta\Gamma_s/\Gamma_s$ measurement.

```
Alias MYJ/psi J/psi
Alias MYphi  phi
Decay B_s0
1.000 MYJ/psi MYphi  SVV_HELAMP 0.70710678 0.0 0.0 0.0 0.70710678 3.1415927;
Enddecay
Decay anti-B_s0
1.000 MYJ/psi MYphi  SVV_HELAMP 0.70710678 3.1415927 0.0 0.0 0.70710678 0.0;
Enddecay
Decay MYJ/psi
1.000 mu+ mu- PHOTOS VLL;
Enddecay
Decay MYphi
1.000 K+ K-          VSS;
Enddecay
End
```

Listing C.5: CP-odd component for p_T cut optimization for $\Delta\Gamma_s/\Gamma_s$ measurement.

```
Alias MYJ/psi J/psi
Alias MYK*0 K*0
Alias MYanti-K*0 anti-K*0
ChargeConj MYK*0 MYanti-K*0
Decay B0
1.000 MYJ/psi MYK*0      SVV_HELAMP 0.160757 1.573832 0.772658 0.0 0.614131 2.710791;
Enddecay
Decay anti-B0
1.000 MYJ/psi MYanti-K*0 SVV_HELAMP 0.614131 2.710791 0.772658 0.0 0.160757 1.573832;
Enddecay
Decay MYJ/psi
1.000 mu+ mu- PHOTOS     VLL;
Enddecay
Decay MYK*0
1.000 pi- K+             VSS;
Enddecay
Decay MYanti-K*0
1.000 pi+ K-             VSS;
Enddecay
End
```

Listing C.6: $B_d \to J/\psi K^{*0}$ sample for the study of the K $\leftrightarrow$ π mass mis-assignment (realistic amplitudes).

Appendix D

Cut optimization

The goal of the cut optimization is to achieve the best measurement possible given the data. We define the "best measurement" as the one with the smallest combined (added in quadrature statistical and systematic) uncertainty on the primary quantity of interest, which is the lifetime in the average lifetime measurement and $\Delta\Gamma_s/\Gamma_s$ in the time-dependent angular analysis of $B_s \to J/\psi\phi$. Below we explain the details and results of the cut optimization procedure as well as the specific considerations we took into account when making the final choice of cuts.

D.1 Cut optimization for the average lifetime measurement

There are two p_T cuts, p_T^B and $p_T^{K/\phi}$ and a vertex quality cut, $Prob(\chi^2)$, that are the major regulators of the composition of the sample in terms of the amount of signal, background and the kinds of backgrounds there are. Let us consider the following abstract exercise involving the actual B_u data sample. Take the sample at some loose cuts, say $p_T^{B_u} > 4.5\,\text{GeV/c}$, $p_T^{K^+} > 1.5\,\text{GeV/c}$ used in skimming and a reasonable vertex quality cut. Let us pretend that the vertex quality cut is fixed by some external conditions and we are only optimizing ($p_T^{B_u}$, $p_T^{K^+}$). What we can do now if we want to extract the B_u lifetime with the smallest statistical uncertainty is to scan the cut space doing the lifetime fit at every pair of cut values. By doing so we find an *optimum*, *i.e.* a pair of cut values that give the best relative statistical uncertainty. We can also get a sense of how much worse we do if we pick a pair of cut values away from the optimum.

However, optimizing on data on which the measurement is to be done is, at best, inappropriate, if the signal in data is less than very large. What might happen is that the optimization latches on a fluctuation (which may be large if the signal is small) and give a result away from optimum. In this section we devise a procedure which avoids this problem by replacing the signal in data with the one from the Monte-Carlo simulation. The fluctuations of the signal are avoided because we substitute the signal in data by as many samples of signal from Monte-Carlo as we want and then average the results. The background fluctuations are not avoided however, because we use the

same background from data over and over again. It should not matter much though, because B meson samples we consider are background dominated and background fluctuations are small.

Technically the cut optimization is implemented in three steps. In step 1 we:

a) apply the nominal cuts (Tables 3.1 and 3.2) on the B_u `NTuple` except that transverse momentum requirements are relaxed to $p_T^{B_u} > 4.5\,\mathrm{GeV/c}$ and $p_T^{K^+} > 1.5\,\mathrm{GeV/c}$ to get the initial sample for cut optimization

b) do a χ^2 fit to the mass histogram from data and extract parameters of signal and background (Figure D-1 (a)),

c) remove from the sample everything but the side-bands, shown with diagonal hatching in Figure D-1 (a),

d) populate the cut-away region by cyclically picking candidates from side-bands and adjusting the mass of the candidate in accordance with the background shape determined in b) above as illustrated by Figure D-1 (b). Get the `NTuple` with background "from data" as the result.

In step 2 we:

a) re-weight the Monte-Carlo sample (Sample 1 described in Appendix C) according to the fitted signal shape from data determined in step 1.b. The re-weighting sequence is illustrated by Figures D-1 (c)–(e).

b) add re-weighted Monte-Carlo signal to the `NTuple` with the background "from data".

We get the optimization `NTuple` as the result (Figure D-1 (f)). As the third and the last step we repeat the following sequence on the optimization `NTuple`:

a) apply a set of (successively tighter) p_T cuts,

b) perform unbinned maximum likelihood fit and record $\sigma_{c\tau}/c\tau$.

Cuts that give the minimum relative $\sigma_{c\tau}$ are the optimal cuts as far as statistical uncertainty of the measurement is concerned.

As the result of the above-described multi-step procedure we have established that more signal (even at the expense of higher background) benefits the statistical power of the measurement, which Table D.1 below confirms. Having recorded the signal yields, N_{sig}, at each pair of cuts, we have established that the scaling of the relative uncertainty on the lifetime approximately follows $0.95/\sqrt{N_{sig}}$.

An interesting aside is that we can perform conventional $S^2/(S+B)$ optimization using the same optimization `NTuple`, by just recording this quantity at each pair of cuts. We have done so and, curiously, this optimization leads to the same results – looser cuts are better – at least in the cut range we have considered. We have also considered yet another optimization technique, based on the speculation that in the

lifetime measurement only the long-lived background matters. In this technique the quantity $S^2/(S+B)$ is maximized only taking into account the events (both signal and background) that satisfy $L_{xy} > 100\,\mu$m. It is rumored to have been used in the CDF Run I lifetime analyses, however we were unable to find an explanation how the optimum found with this technique translates into the best statistical uncertainty. This optimization returned a different and, consequently, non-optimal result ($p_T^{B_u} > 6.5\,\mathrm{GeV/c}^2$ and $p_T^{K^+} > 2.0\,\mathrm{GeV/c}^2$). We suspect that what plays a role is the fact that we do a simultaneous MLL fit to the mass and the PDL of *all* candidates, and the signal with $L_{xy} < 100\,\mu$m contributes to the measurement non-trivially.

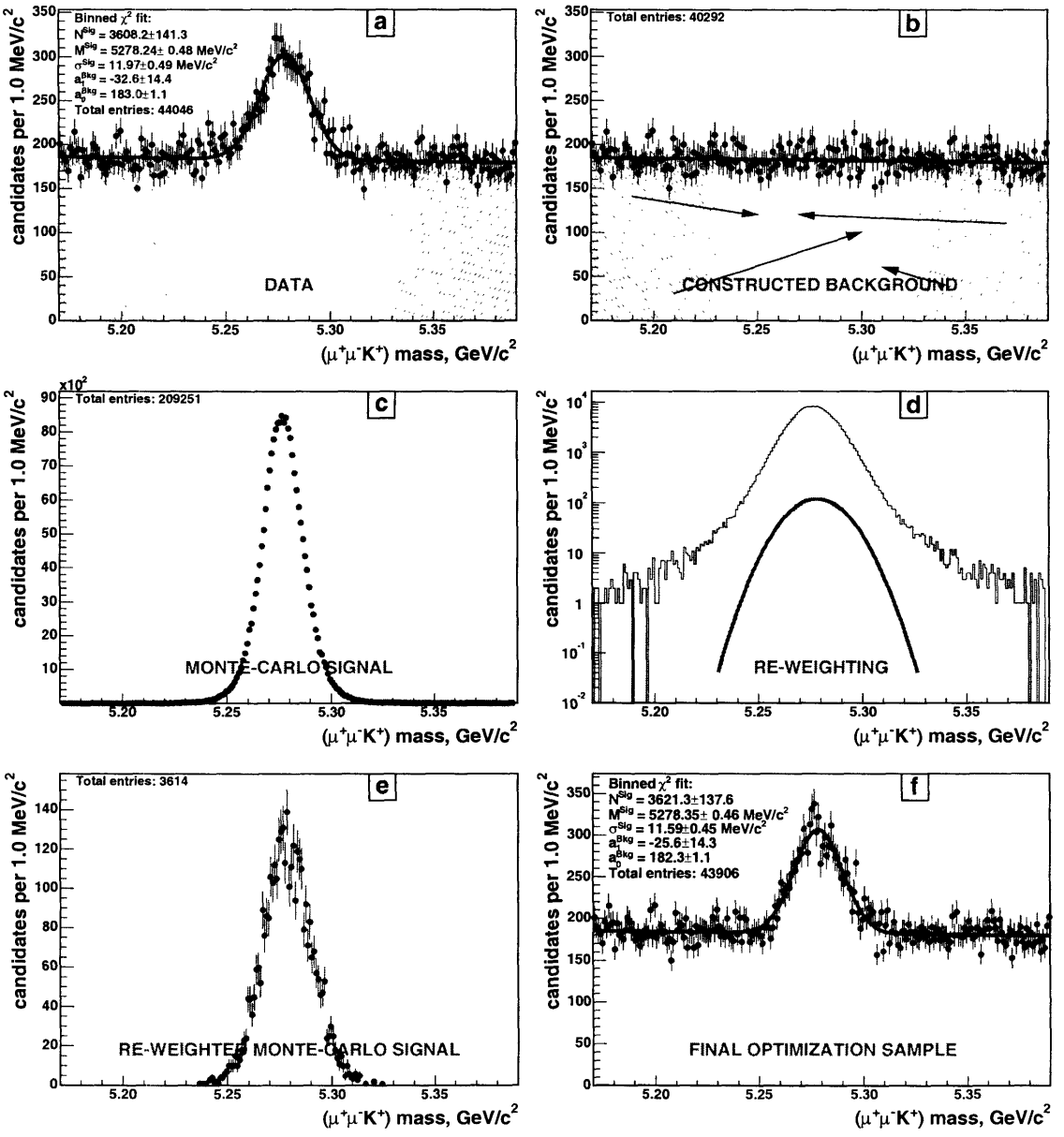


Figure D-1: Illustration of the cut optimization procedure. See text for the explanation.

Another conclusion that we make as the result of performing the cut optimization procedure is that the optimum is rather broad and if one steps away from it the penalty in statistical uncertainty is not severe.

	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3
4.5	1.8287	1.8560	1.8781	1.9138	1.9428	1.9775	2.0234	2.0649	2.1069
5.0	1.8532	1.8795	1.9000	1.9325	1.9588	1.9912	2.0339	2.0738	2.1157
5.5	1.8905	1.9176	1.9386	1.9716	1.9964	2.0269	2.0706	2.1060	2.1424
6.0	1.9301	1.9558	1.9762	2.0104	2.0352	2.0637	2.1070	2.1430	2.1780
6.5	1.9881	2.0105	2.0310	2.0643	2.0861	2.1162	2.1591	2.1935	2.2261
7.0	2.0438	2.0671	2.0858	2.1165	2.1385	2.1667	2.2093	2.2448	2.2791
7.5	2.1134	2.1349	2.1523	2.1840	2.2033	2.2272	2.2705	2.3033	2.3388

Table D.1: Relative uncertainty (in percent) as a function of $p_T^{B_u}$ (rows) and $p_T^{K^+}$ (columns).

D.2 Other considerations in the choice of cuts

In the preceding section we describe the procedure of optimizing the p_T cuts with respect to *statistical* uncertainty of the extracted lifetime. The procedure is unbiased, *i.e.* free of subjective input or tuning, and efficient, because it optimizes directly on the uncertainty of the quantity of interest. However, there are at least three important considerations which bar one from setting the cuts entirely relying just on the results of the optimization procedure.

The first two are somewhat entangled. They have to do with the amount of badly mis-reconstructed events and the fit model used being appropriate for one set of cuts while being inappropriate for another set of cuts. Figures D-2 and D-3 illustrate the discussion that follows. Figure D-2 shows B_d mass and proper decay length distributions at relatively loose cuts. It is obvious that the nominal model needs to be updated to include a description of the long-lived negative tail if we were to fit that particular sample. If the model is not amended then one is to pay a price in systematic uncertainty if the fit converges at all. In Figure D-3 we see another extreme – fairly tight cuts. There does not seem to be much of the negatively lived tail, which indicates that one might have to fix the corresponding contribution in the model to zero in order to make the fit converge. There is a potential gain in systematic uncertainty due to somewhat less ambiguity in the choice of the model, but there is a price to pay in statistical uncertainty when tighter cuts are used.

The problem with the badly mis-reconstructed events (such as long-lived negative tail and its positively lived partner) is that the number of them does not necessarily have to scale with tightness of "physics" cuts for the very reason that these are "unphysical" events. Just to elaborate on this point: for example knowing how much signal we have at one set of p_T cuts we can predict how much signal we will have at some other p_T cuts because we know the p_T spectra involved. Same holds true,

at least qualitatively, for the other components in the sample, not just signal, but not for the badly mis-reconstructed events. It might take really tight cuts, which would greatly reduce the statistical precision of the measurement, to get rid of the mis-reconstructed events completely, because, say one or two of them happened to have fairly high p_T. So the best possibility would be to find some sort of "magic bullet" – a cut (not necessarily a "physics" cut) that would eliminate most of the mis-reconstructed events, without hurting the signal too much or introducing any kind of bias. We took on such a quest and it turned out to be at least partially successful (see Appendix F).

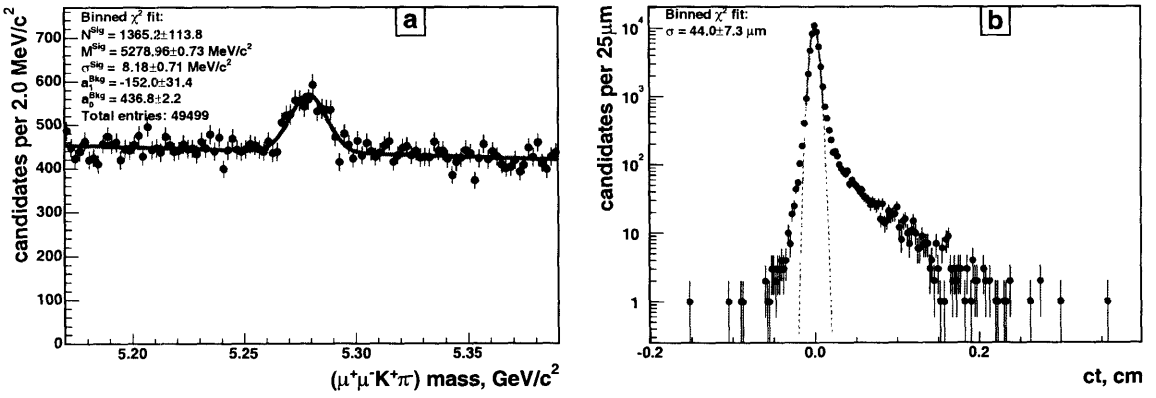


Figure D-2: Distribution of (a) mass and (b) PDL of the $B_d \to J/\psi K^{*0}$ candidates reconstructed with $p_T^{B_d} > 4.5\,\text{GeV/c}$ and $p_T^{K^{*0}} > 1.5\,\text{GeV/c}$. The PDL distribution is fit to a Gaussian in the range from $-100\,\mu$m to $100\,\mu$m. The fit is extrapolated outside that region "to guide the eye".

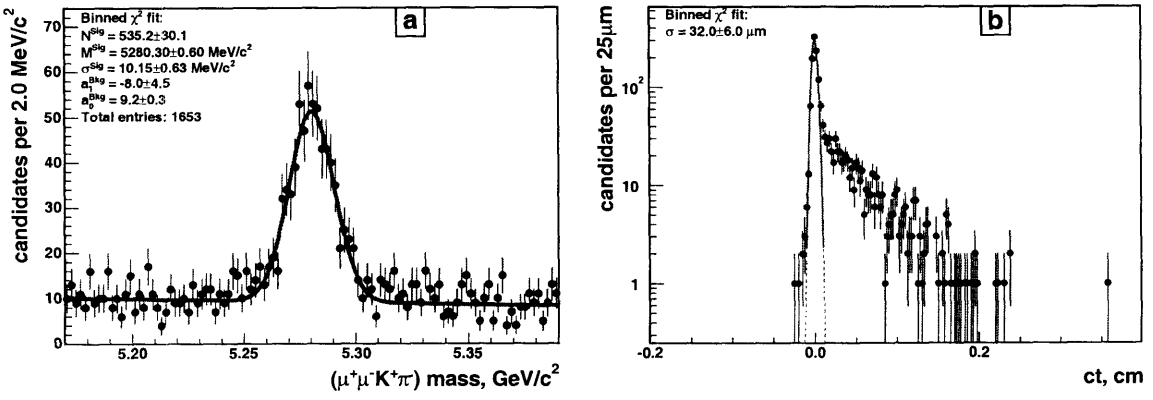


Figure D-3: Distribution of (a) mass and (b) PDL of the $B_d \to J/\psi K^{*0}$ candidates reconstructed with $p_T^{B_d} > 10.0\,\text{GeV/c}$ and $p_T^{K^{*0}} > 3.5\,\text{GeV/c}$. The PDL distribution is fit to a Gaussian in the range from $-100\,\mu$m to $100\,\mu$m. The fit is extrapolated outside that region "to guide the eye".

In order to decide if there are mis-reconstructed events, if the short-lived negative tail is present, *etc.* one needs to look at the data, at least at the PDL distribution.

There is a danger involved — that of subjectivity. Some subjectivity is inevitable, but we try to minimize it with the cut choice procedure that we describe at the end of this section, right after we introduce the last of the three considerations for the choice of cuts.

This last consideration comes from the fact that the ultimate goal of the analysis is to measure $\Delta\Gamma/\Gamma$ in the B_s sample, which requires a separation of the CP eigenmodes. Reference [106] argues, that the angular analysis needed to achieve such separation loses at least some of its power in the presence of large backgrounds, while large backgrounds (as long as they are well-behaved) do not seem to be a problem for the lifetime measurements. What this means is that at least in neutral B samples tighter cuts will not necessarily be bad.

Let us now turn to how we chose the numerical values for the cuts to be used in the analysis. First, we would like to review the vertex quality cut.

One would think that true vertices, such as those of signal and prompt background, have flat $Prob(\chi^2)$ distribution and badly mis-reconstructed vertices would cluster at low probabilities. It is, unfortunately, not the case — $Prob(\chi^2)$ peaks at low values for the signal too, as one can see from Figure D-4. This issue is something that still needs to be understood, but not necessarily in the context of the current analyses. Optimizing a cut that is not quite understood is a bad option, however we do not want to ignore the utility of the $Prob(\chi^2)$ in rejecting at least some very unlikely (very low quality) vertices.

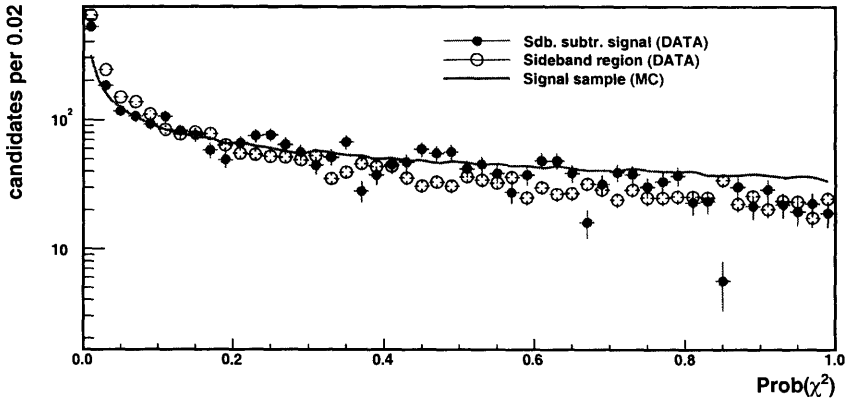


Figure D-4: Vertex quality ($Prob(\chi^2)$) for B_u candidates with nominal selection and the $Prob(\chi^2) > 10^{-3}$ cut applied. All plots are normalized to the area of the side-band subtracted signal from data.

In a range of values from 10^{-6} to 10^{-3} the vertex quality cut affects the size of the signal rather insignificantly — the variation is about 8% when going from one of the above extremes to the other as tested on B_u sample. Also, it seems, that once the vertex quality is at $10^{-5} \ldots 10^{-3}$ cutting tighter on it in attempt to reduce the contribution of badly mis-reconstructed events is not more beneficial than raising p_T thresholds. In order to avoid "tuning on data", at least to some extent, we limit ourselves to picking the numerical value for this cut from 10^{-3}, 10^{-4}, 10^{-5}, ... series. Below we list the choices we have made along with the reasoning we have exercised:

- $B_u \to J/\psi K^+$: $Prob(\chi^2) > 10^{-3}$ — 3 track vertex
- $B_d \to J/\psi K^{*0}$: $Prob(\chi^2) > 10^{-4}$ — 4 track (better determined) vertex
- $B_s \to J/\psi\phi$: $Prob(\chi^2) > 10^{-5}$ — 4 track vertex and low combinatorics due to narrow ϕ mass window

The next step is to chose p_T cuts. The optimization procedure tells that we want them to be as loose as possible, however we do not want to have any significant contribution from badly mis-reconstructed events in our sample. We have decided to choose the loosest possible cuts that reject all candidates with $ct < -500\,\mu$m; however, we increment p_T^B cut only in 0.5 GeV/c steps[1] and $p_T^{K/\phi}$ only in 0.1 GeV/c steps[2]. The resulting cuts are listed in Table 3.2.

D.3 P_T cut optimization for $\Delta\Gamma_s/\Gamma_s$ measurement

The cut optimization procedure for $\Delta\Gamma_s/\Gamma_s$ is very similar to what is described in Section D.1 for the average lifetime measurement. The only difference/complication is that the signal now consists of the two components: CP-even and CP-odd.

We use Monte-Carlo Sample 4 and 5 (Appendix C) to replace the signal in the data sample. Before using them, however, we perform the mass-lifetimes-amplitudes fit (excluding the background component) on the combination of these two samples to make sure that the generation was done successfully. The results of this exercise are summarized in Table D.2.

Par.	Result of the fit	Input	Unit
M	5368.63 ± 0.02	5369.60	MeV/c^2
S_m	1.876 ± 0.004	—	
$\lvert A_0\rvert^2$	0.5627 ± 0.0019	0.5625	
$\lvert A_\parallel\rvert^2$	0.2048 ± 0.0029	0.2025	
$\delta_\parallel$	1.980 ± 0.014	2.0	
$c\tau_L$	332.1 ± 1.4	330.0	μm
$\Delta\Gamma/\Gamma$	49.6 ± 0.9	50.0	%
N_{sig}	132129		

Table D.2: Results of the full mass-lifetimes-amplitudes fit on the Monte-Carlo Samples 4 and 5 used in p_T cut optimization.

Instead of just several optimization `NTuples` we prepare 24 of them ensuring that the signal components in them are statistically independent from each other. The increase in the number of the `NTuples` is to compensate for a greater chance of getting a sample with a large fluctuation, when the signal size is only about 200 events

[1] Starting from $p_T^B = 4.5$ GeV/c, which is the cut applied when producing the skimmed samples.

[2] Starting from $p_T^{K/\phi} = 1.5$ GeV/c, which is the cut applied when producing the skimmed samples.

and the signal itself is not homogeneous but breaks down into CP-odd and CP-even components.

We look for an optimal cut combination among 16 pairs, which are assigned a number according to Table D.3.

	1.5	1.7	1.9	2.1
5.0	1	5	9	13
5.5	2	6	10	14
6.0	3	7	11	15
6.5	4	8	12	16

Table D.3: Optimization grid: rows – p_T^{ϕ} ($p_T^{B_s}$ constant), columns – $p_T^{B_s}$ (p_T^{ϕ} constant). Each cell assigns an integer number to the cut combination given by cell's coordinates. This integer number is then used to denote the cut combination throughout the optimization procedure.

Figure D-5 illustrates the optimization procedure as it plots the absolute uncertainty on $\Delta\Gamma_s/\Gamma_s$ obtained from each of the 24 optimization `NTuples` as a function of cuts used. It shows no strong dependence of the absolute uncertainty of the $\Delta\Gamma_s/\Gamma_s$

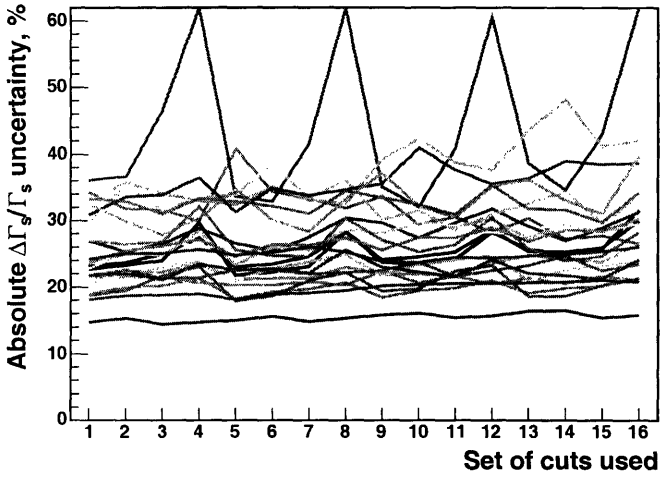


Figure D-5: Absolute uncertainty on $\Delta\Gamma_s/\Gamma_s$ as a function of cuts used. Each line represents one of the 24 optimization `NTuples`.

measurement on the p_T cuts used. Perhaps, one should avoid $p_T^{B_s} > 6.5\,\text{GeV/c}$ and tighter cuts as the structure in Figure D-5 suggests. Other than this, cuts can be chosen based on other considerations.

In this analysis one of the important cross-checks is based on the CP-odd content measurement with mass-and-amplitudes-only fit performed for different ct cut values. Such measurements benefit from lower background level which is achieved by tighter p_T cuts. We therefore adopt $p_T^{B_s} > 6.0\,\text{GeV/c}$ and $p_T^{\phi} > 2.0\,\text{GeV/c}$ as our nominal B_s time-dependent angular analysis requirements.

Appendix E

$B_d \to J/\psi K^{*0}$ and $B_s \to J/\psi\phi$ cross-feed

In this appendix we estimate the amount of cross-feed between $B_d \to J/\psi K^{*0}$ and $B_s \to J/\psi\phi$. At our disposal are the two Monte-Carlo samples: $B_d \to J/\psi K^{*0}$ (80M events generated) and $B_s \to J/\psi\phi$ (40M events generated), which are correspondingly Sample 2 and 3 from Appendix C. Both are generated using `EvtGen`'s `PHSP` decay model (no polarization, just the phase space decay), but the relative detector acceptance is unlikely to affect the semi-quantitative conclusions we make here, especially given that the polarizations in the two decays are similar.

We first estimate the relative frequency with which the two decays considered occur in nature. Using numbers from Reference [25] we obtain:

$$\frac{f_d}{f_s} \cdot \frac{Br(B_d \to J/\psi K^{*0}) \times Br(K^{*0} \to K^+\pi^-)}{Br(B_s \to J/\psi\phi) \times Br(\phi \to K^+K^-)} = \frac{38.8\%}{10.6\%} \cdot \frac{1.31 \cdot 10^{-3} \times 0.6657}{0.93 \cdot 10^{-3} \times 0.492} \approx 7, \tag{E.1}$$

which means that we have to scale the results from $B_d \to J/\psi K^{*0}$ sample by 7/2 in order to make meaningful comparisons to the $B_s \to J/\psi\phi$ sample (2 =80M/40M). We apply this scaling in all plots included in this appendix.

As the first exercise we run B_s reconstruction code and apply the "nominal" B_s selection on both of the above-mentioned Monte-Carlo samples[1]. The resulting K^+K^- and $\mu^+\mu^-K^+K^-$ mass distributions are shown in Figure E-1. From these plots it is evident that with ϕ mass window of $\pm 10\,\mathrm{MeV/c^2}$ around M^{ϕ}_{PDG} there is about 4% contamination in the B_s sample coming from B_d. The only way to reduce this contamination (and the associated systematic uncertainty) is to narrow down the ϕ mass window. It is also time to review the size of this window for a different reason: with much improved tracking the detector mass resolution is improved potentially making $\pm 10\,\mathrm{MeV/c^2}$ window used in Run I and early in Run II unnecessarily too

[1]The word "nominal" is taken in quotes because, as it turns out, we are defining part of what "nominal" is in this section. However, we use the same technical quality cuts throughout.

wide.

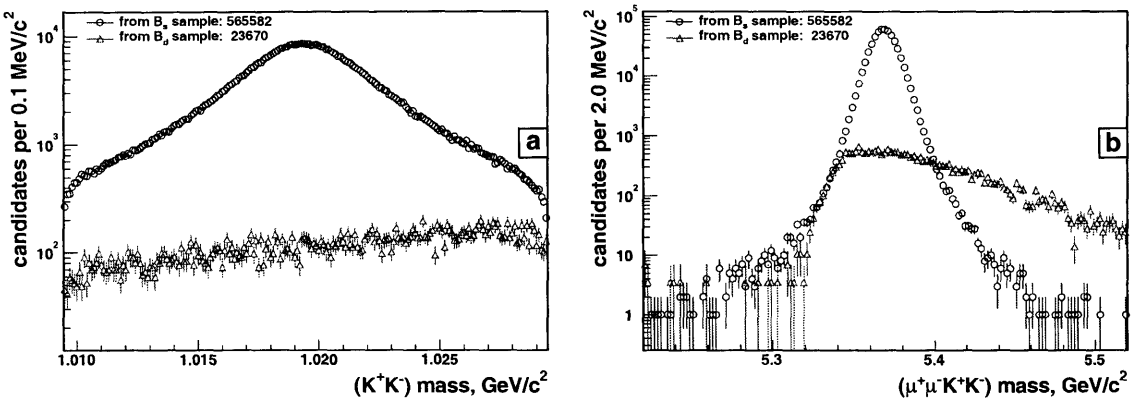


Figure E-1: Mass distributions and yields obtained when applying B_s selection with a $\pm 10\,\mathrm{MeV/c^2}$ ϕ mass window to the (properly normalized) samples of $B_d \to J/\psi K^{*0}$ and $B_s \to J/\psi\phi$ decays.

To figure out the appropriate size for the mass window around M^{ϕ}_{PDG} we plot the distribution of the K^+K^- mass obtained from `xpmm0c` dataset; track quality requirements specified in Table 3.1 are applied. In fact, we make plots for four values of the cut on p_T of ϕ candidates (1.6 GeV/c, 2.0 GeV/c, 4.0 GeV/c and 10.0 GeV/c), anticipating that the width of the ϕ peak may depend on p_T. These plots are shown in Figure E-2. Each distribution is fit to a Gaussian signal sitting atop a quadratic background. The result of the fit is shown in each plot.

We observe that the width of the ϕ peak is about $2.5 - 2.6\,\mathrm{MeV/c^2}$ and there is no strong p_T dependence (largely due to the low Q-value in $\phi \to K^+K^-$ decay). The extracted mass of the ϕ resonance agrees with the PDG value within the systematic uncertainty of the momentum scale calibration [94]. We then set a *new* size of the mass window for the ϕ candidates at $\pm 6.5\,\mathrm{MeV/c^2}$ ($\pm 2.5\sigma$, $\sigma = 2.6\,\mathrm{MeV/c^2}$). It reduces the B_d contamination of the B_s sample from 4.02% to about 2.85% as can be derived by contrasting numbers in Figure E-3 and Figure E-1.

A legitimate question is how what we gain in systematic uncertainty compares to what we lose in statistical uncertainty. Based on Monte-Carlo numbers we expect the $\pm 6.5\,\mathrm{MeV/c^2}$ mass window to be 91% efficient (4.5% increase in stat. uncertainty) for the signal, compared to the $\pm 10.0\,\mathrm{MeV/c^2}$ window. The statistical uncertainty with $\pm 10.0\,\mathrm{MeV/c^2}$ mass window is about $28\,\mu$m, so narrowing the window down to $\pm 6.5\,\mathrm{MeV/c^2}$ would increase the statistical uncertainty by $\sim 1.3\,\mu$m. The systematic uncertainty on the average lifetime will (naively) go down by approximately $(4.02 - 2.85) \cdot 10^{-2} \times (c\tau_{B_d} - c\tau_{B_s}) \approx 1.2 \cdot 10^{-2} \cdot 60\,\mu\mathrm{m} \approx 0.7\,\mu$m. It looks like at present level of statistical and systematic uncertainties it is beneficial to stay with the larger mass window, however we go for the smaller one because of the following reasons:

- statistical uncertainty will shrink as more data are accumulated, systematic uncertainty due to contamination is not going to change, so the narrower mass window will only work better in the future

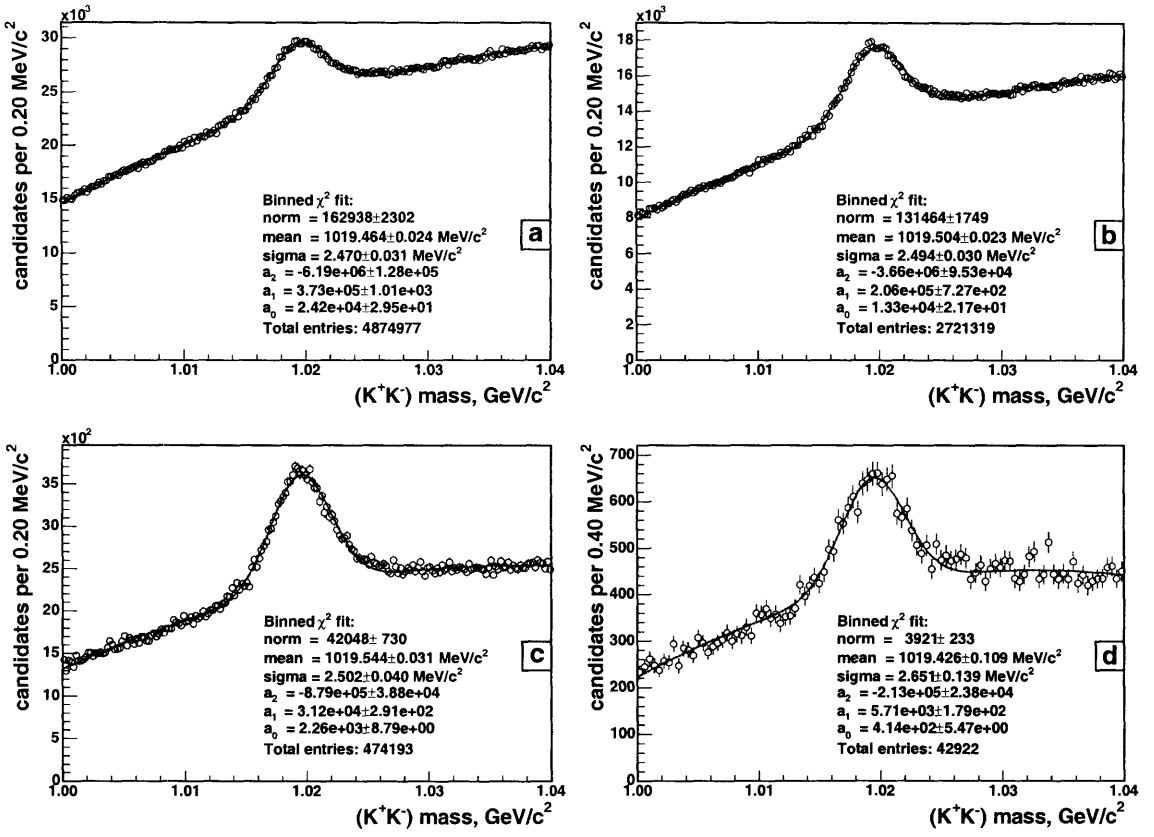


Figure E-2: Mass distribution of ϕ (K^+K^-) candidates from the `xpmm0c` data set with the requirement of: (a) $p_T^\phi > 1.6\,\mathrm{GeV/c}$, (b) $p_T^\phi > 2.0\,\mathrm{GeV/c}$, (c) $p_T^\phi > 4.0\,\mathrm{GeV/c}$, (d) $p_T^\phi > 10.0\,\mathrm{GeV/c}$.

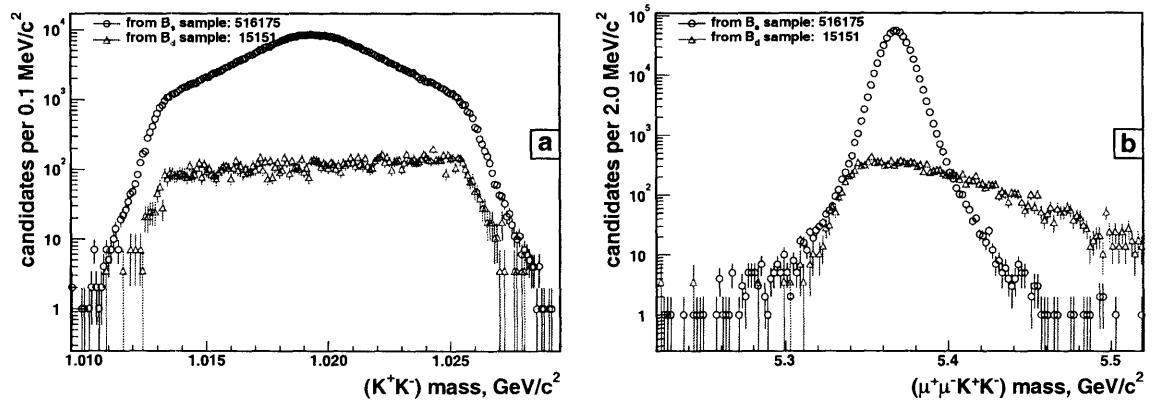


Figure E-3: Mass distributions and yields obtained when applying B_s selection with a $\pm 6.5\,\mathrm{MeV/c^2}$ ϕ mass window to the (properly normalized) samples of $B_d \to J/\psi K^{*0}$ and $B_s \to J/\psi\phi$ decays.

- B_d contamination seems to be the largest source of systematic uncertainty (both for the average lifetime and the $\Delta\Gamma/\Gamma$ measurements) which we reduce by 29% with the narrower window

- narrower window reduces combinatorial background in data by more than 1/3, which helps to separate different CP components in the angular analysis.

Now let us investigate the inverse of the above problem: how much contamination is there in the $B_d \to J/\psi K^{*0}$ sample because of $B_s \to J/\psi\phi$ decays. To understand this issue we run B_d reconstruction code and apply the "nominal" B_d selection, in particular K^{*0} mass window of $\pm 80\,\mathrm{MeV/c^2}$ used in Run I and early in Run II, on both of the Monte-Carlo samples. The resulting $K^+\pi^-$ and $\mu^+\mu^- K^+\pi^-$ mass distributions are shown in Figure E-4.

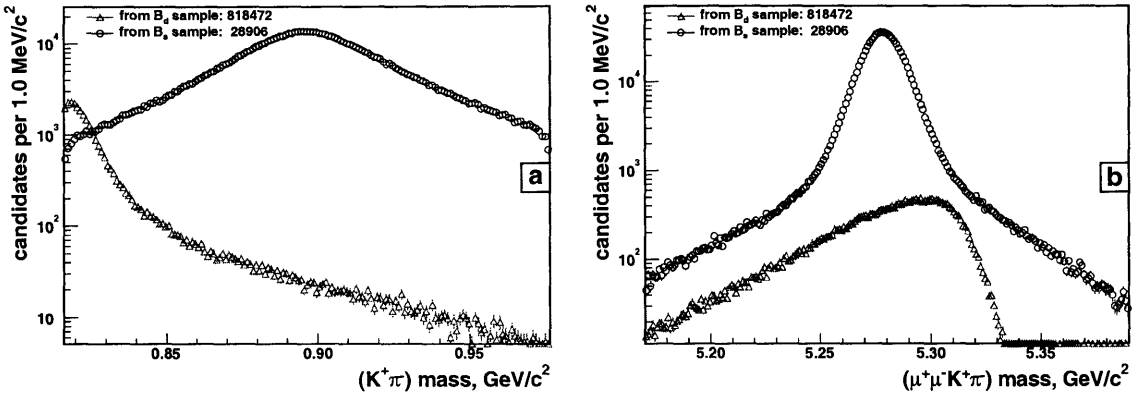


Figure E-4: Mass distributions and yields obtained when applying B_d selection with a $\pm 80\,\mathrm{MeV/c^2}$ K^{*0} mass window to the (properly normalized) samples of $B_d \to J/\psi K^{*0}$ and $B_s \to J/\psi\phi$ decays.

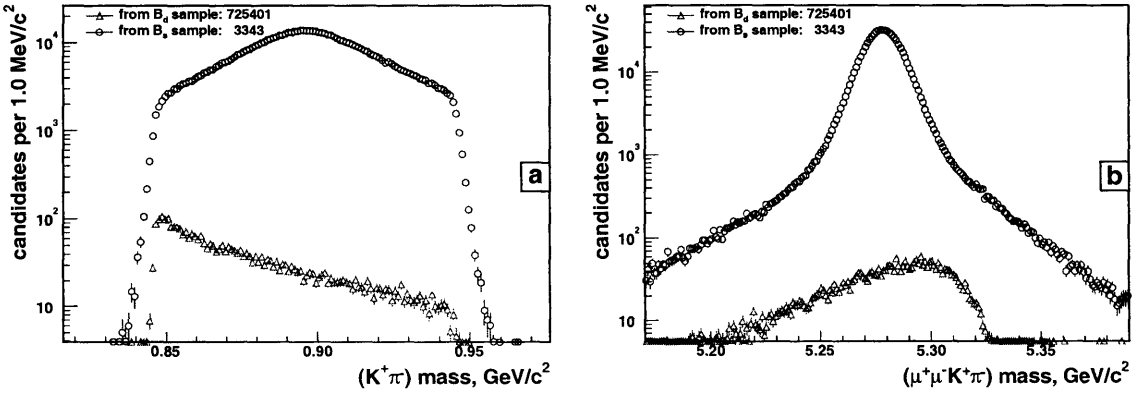


Figure E-5: Mass distributions and yields obtained when applying B_d selection with a $\pm 50\,\mathrm{MeV/c^2}$ K^{*0} mass window to the (properly normalized) samples of $B_d \to J/\psi K^{*0}$ and $B_s \to J/\psi\phi$ decays.

B_s contamination is about 3.41% and most of it comes from the $K^+\pi^-$ combinations with the invariant mass below 840 MeV/c^2, as has been established earlier in Reference [103]. Rather than simply rejecting all $K^+\pi^-$ candidates with mass below 840 MeV/c^2, we choose to adjust the K^{*0} mass window symmetrically to 896.1 MeV/c^2 $\pm$ 50 MeV/c^2 (see Figure E-5). This requirement is 89% efficient on signal (stat. uncertainty 14.1 μm $\rightarrow$ 15.0 μm) while reducing the contamination to 0.45% (syst. uncertainty due to B_s contamination 1.9 μm $\rightarrow$ 0.3 μm). An additional benefit is that the combinatorial background (assuming linear shape under the K^{*0} peak) is reduced by about 38%, which is helpful in the angular analysis [106].

Appendix F

Eliminating events with large negative *ct*

We have studied numerous possibilities that at least in theory could lead to eliminating the badly mis-reconstructed events without uncontrollable damage to the statistics of the measurement. We have found partial success by looking at the origin of the J/ψ candidates with pseudo-*ct*, defined as $M_{PDG}^{J/\psi} \cdot \frac{L_{xy}}{p_T}$, less than $-500\,\mu$m.

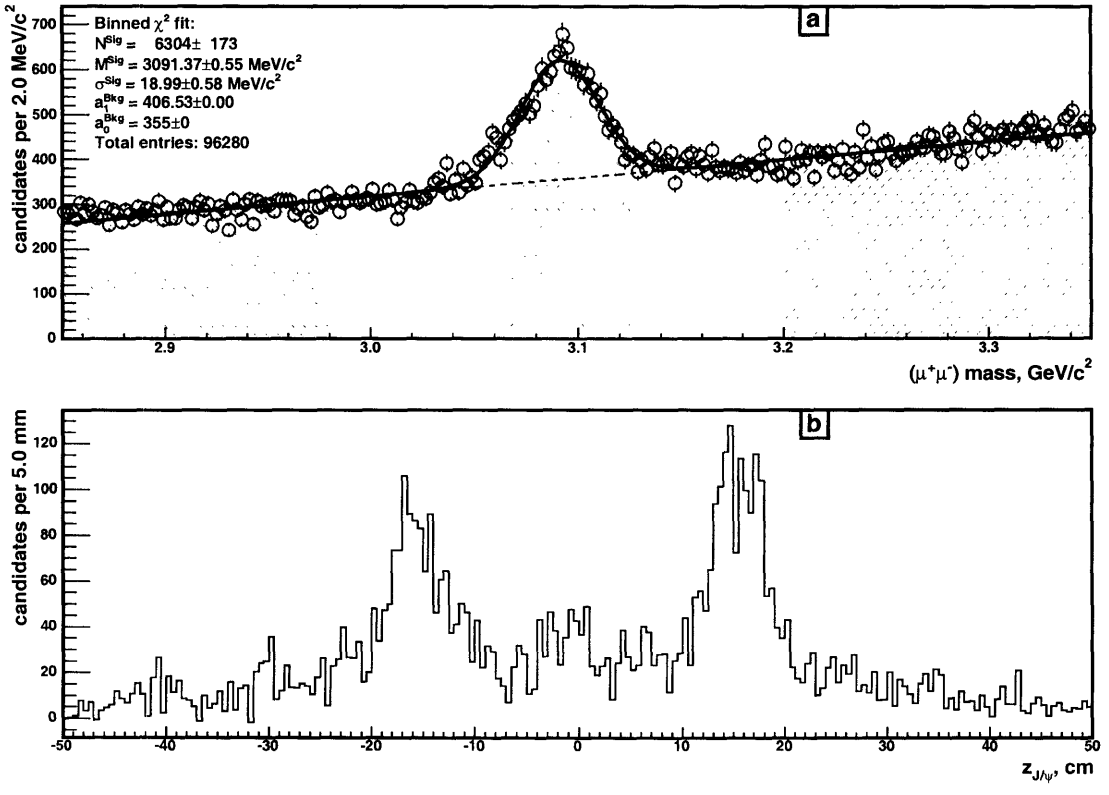


Figure F-1: (a) Mass distribution of the J/ψ candidates with $ct_{pseudo} < -500\,\mu$m. This plot defines side-bands (diagonal hatching) and signal region (cross-hatching). (b) Side-band subtracted distribution of z positions of the J/ψ vertices.

Plot in Figure F-1 (b) shows side-band subtracted distribution of the vertex z position of such J/ψ candidates. It clearly indicates that long negatively lived tail is dominated by the bulkhead region. This plot is to be contrasted with the same distribution for all J/ψ candidates, passing technical quality cuts, shown in Figure F-2.

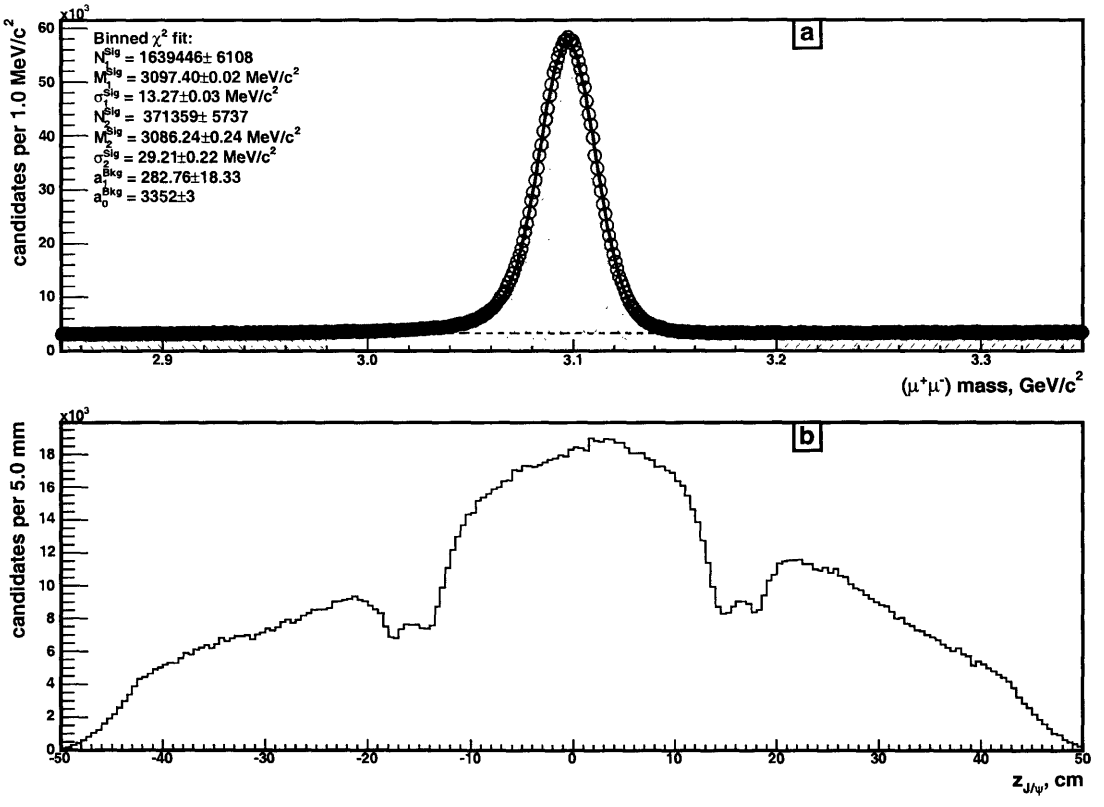


Figure F-2: (a) Mass distribution of all J/ψ candidates. Side-bands (diagonal hatching) and signal region (cross-hatching) are shown. (b) Side-band subtracted distribution of z positions of the J/ψ vertices.

Excluding B candidates with z position of the vertex falling into $(-18.0\,\text{cm}; -14.0\,\text{cm})$ or $(14.5\,\text{cm}; 18.5\,\text{cm})$ intervals helps to greatly reduce the amount of badly misreconstructed events, even at the loosest "physics" cuts.

Appendix G

Calculating convolutions

It is always nice to see a mathematical expression that can be easily converted into a program code. The convolution operation, $\otimes$, is not one of the intrinsic functions in any major programming language, therefore it is often needed that convolution $A \otimes B$ is calculated analytically before it is used for computation in any sort of computer code. One might exercise an option of using `MATHEMATICA`'s Integrator[1], but it is not always helpful. Below we calculate two convolution integrals used in this dissertation in terms of functions implemented in `ROOT` [93].

Equation G.1 is what we use to convolute positively lived exponentials, such as the signal and (f_+, λ_+), (f_{++}, λ_{++}) components of the background, with a Gaussian resolution function:

$$\begin{aligned} E(x|\lambda) \otimes G(x, \sigma_x|S) &= \int_{-\infty}^{\infty} dx' \cdot E(x'|\lambda) \cdot G(x - x', \sigma_x|S) \\ &= \int_{0}^{\infty} dx' \cdot \frac{1}{\lambda} e^{-\frac{x'}{\lambda}} \cdot \frac{1}{\sqrt{2\pi} S \sigma_x} e^{-\frac{(x-x')^2}{2S^2\sigma_x^2}} \\ &= \frac{1}{2\lambda} e^{\frac{S^2\sigma_x^2}{2\lambda^2} - \frac{x}{\lambda}} \left[1 - \mathrm{Erf}\Big(\frac{S\sigma_x}{\sqrt{2}\lambda} - \frac{x}{\sqrt{2} S \sigma_x} \Big) \right], \end{aligned} \tag{G.1}$$

where Erf() is available in `ROOT` as `TMath::Erf()`. In order to obtain the result of a convolution for the negatively lived exponential, such as (f_-, λ_-), one needs just to flip the sign of x before plugging it in this formula.

When considering a resolution function with symmetric exponential tails, as we do in Section 3.4.3, the following seemingly trivial convolution arises:

$$E(x|\lambda) \otimes e^{-\frac{|x|}{\lambda_R}} = \int_{-\infty}^{\infty} dx' \cdot E(x'|\lambda) \cdot e^{-\frac{|x-x'|}{\lambda_R}}, \tag{G.2}$$

[1] *http://www.integrals.com/*

where $E(x|\lambda)$ is defined in Equation 3.5. It turns out not so trivial when it comes to the actual calculation of the integral which first breaks down into two cases: $\lambda = \lambda_R$ and $\lambda \neq \lambda_R$, each of which then splits into $x < x'$ and $x \geqslant x'$ case. Moreover, in some of the above cases situation of $x < 0$ has to be treated separately because of the shape of $E(x|\lambda)$. Sparing the reader of the algebra, we present just the final result of the calculation:

$$E(x|\lambda) \otimes e^{-\frac{|x|}{\lambda_R}} = \begin{cases} \frac{1}{2(\lambda+\lambda_R)} e^{\frac{x}{\lambda_R}} & , \; x < 0 \\ \frac{1}{2(\lambda-\lambda_R)} \left[e^{-\frac{x}{\lambda}} - e^{-\frac{x}{\lambda_R}} \right] & , \quad \lambda \neq \lambda_R, \; x \geqslant 0 \\ \left[\frac{x}{2\lambda^2} + \frac{1}{4\lambda} \right] e^{-\frac{x}{\lambda}} & , \quad \lambda = \lambda_R, \; x \geqslant 0 \end{cases} \tag{G.3}$$

Appendix H

Treating $B_u \to J/\psi\pi^+$ in B_u sample

Reference [25] quotes the branching ratio of $B_u \to J/\psi\pi^+$ at $(4.0 \pm 0.5) \cdot 10^{-5}$. The same source gives $(1.00 \pm 0.04) \cdot 10^{-3}$ as the branching ratio for $B_u \to J/\psi K^+$, so in principle if we assign a π^+ from $B_u \to J/\psi\pi^+$ the kaon mass we will have about 4% additional contribution to our sample of $B_u \to J/\psi K^+$. This will be a confusing contribution in a sense that it will distort the mass distribution and put candidates with the lifetime close to that of our signal where we do not want them, more precisely, where the simple fit model does not expect them. But we can turn this around and use such reconstructed $B_u \to J/\psi\pi^+$ decays to measure the lifetime of our signal.

Figure H-1 shows the mass and the proper decay length distributions of the $B_u \to J/\psi K^+$ candidates reconstructed in the Monte-Carlo sample of $B_u \to J/\psi\pi^+$ decays. The mass shape is parameterized using the following PDF:

$$X^{obs}_{sig,J/\psi\pi^+}(m) = \begin{cases} 0 & , \quad m < A \\ \frac{1}{N}(m-A)^6 \cdot e^{-\frac{m-A}{B}} & , \quad A \leqslant m \leqslant M^{max} \\ 0 & , \quad m > M^{max} \end{cases} \tag{H.1}$$

where the normalization factor is given by

$$\begin{aligned} N = 720B^7 - e^{-\frac{M^{max}-A}{B}} \times \{ & 720B^7 + \\ & 720B^6(M^{max}-A) + \\ & 360B^5(M^{max}-A)^2 + \\ & 120B^4(M^{max}-A)^3 + \\ & 30B^3(M^{max}-A)^4 + \\ & 5B^2(M^{max}-A)^5 + \\ & B\,(M^{max}-A)^6\}. \end{aligned} \tag{H.2}$$

A and B are parameters extracted by fitting the distribution (values can be read off Figure H-1 (a)), $M^{max} = 5.39\,\mathrm{GeV/c^2}$ is the upper mass window edge.

In order to determine the effective lifetime of the $B_u \to J/\psi\pi^+$ candidates recon-

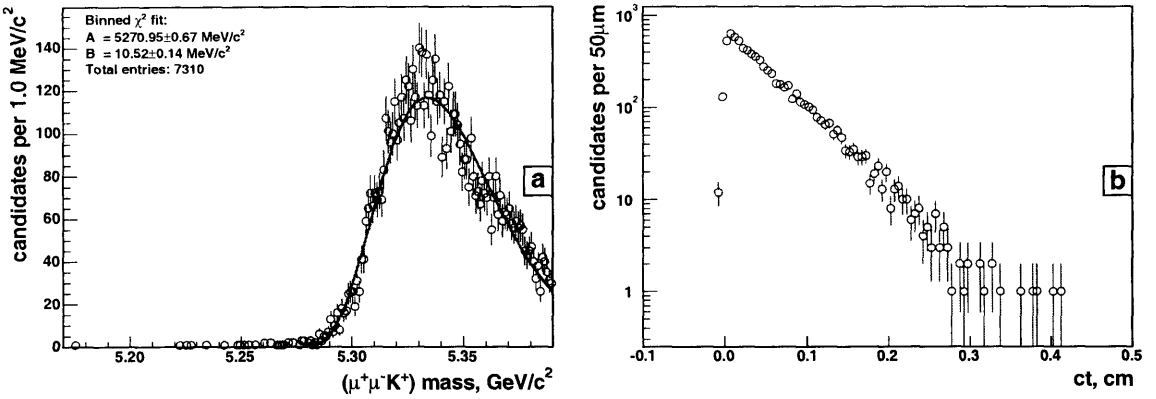


Figure H-1: Mass (a) and proper decay length (b) distributions of the $B_u \to J/\psi K^+$ candidates reconstructed in the Monte-Carlo sample of $B_u \to J/\psi\pi^+$ decays. The mass distribution is fit to $(m-A)^6 \cdot e^{-\frac{m-A}{B}}$. A fit to the PDL of these candidates is discussed in the text.

structed as $B_u \to J/\psi K^+$, we fit their PDLs to an exponential convoluted with a Gaussian in an unbinned maximum likelihood fit. The extracted lifetime is (502.32 ± 5.49) μm, which is to be compared to (497.38 ± 1.01) μm obtained from a $B_u \to J/\psi K^+$ Monte-Carlo sample (Table 3.8). Based on this comparison, in the fit on the actual data sample we fix the lifetime of the $B_u \to J/\psi\pi^+$ contribution to that of our signal times $1.01(\approx 502.32/497.38)$.

Finally, in Reference [97] the relative detector/reconstruction efficiency of the $B_u \to J/\psi\pi^+$ to $B_u \to J/\psi K^+$ is estimated at $99.1 \pm 0.8\%$. We take this number to be 1. In the nominal fit on data we apply a Gaussian constraint on the ratio of the two branching fractions, using the numbers and uncertainties mentioned at the very beginning of this appendix.

Appendix I

Treating $K \leftrightarrow \pi$ self-reflection in B_d sample

We have generated 15M $B_d \to J/\psi K^{*0}$ decays using angular amplitudes determined in Reference [56]. All these events have been subjected to the detector and trigger simulation, after which nominal selection (Tables 3.1 and 3.2) has been applied. The above-described procedure yielded a sample of about 134K fully reconstructed decays, which is described as Sample 6 in Appendix C. We are looking to compare mass and lifetime distributions of the swapped candidates to their correctly reconstructed counterparts. We are not looking to compare the angular distributions of the two, as this issue is dealt with in Section 3.2.4 of the dissertation. We are assuming that neither mass, nor PDL of a swapped candidate are correlated to where the candidate is in the angular space, as plots from Monte-Carlo simulation reveal no such correlation.

Figure I-1 compares the mass distributions of the candidates reconstructed with the correct $K\pi$ mass assignment to the same for the candidates in which incorrect mass assignment was preferred by the selection algorithm. Large statistics makes evident some amount of mis-reconstruction, *i.e.* long tails in the mass distribution of B_d candidates reconstructed with correct mass assignments to the tracks making up the candidate. However, such effects are small enough to be neglected when dealing with the issue at hand.

A more subtle issue is how to parameterize the self-reflection shape. Figure I-2 contrasts the B_d peak in data with the combined B_d peak in Monte-Carlo, for which we do not use the truth information to infer the (in)correctness of the K^{*0} reconstruction. Both distributions are obtained applying $L_{xy} > 100\,\mu$m cut in addition to the nominal requirements. A single Gaussian is fit to each of the peaks, however in data there is also a linear component to describe the background, while in the Monte-Carlo the fit range is limited to $\pm 20.0\,\mathrm{MeV/c^2}$ around the mean of the peak, so that the extracted width is not strongly affected by the mis-reconstructed tails. We attribute the visible width difference to an additional detector smearing which is not modeled in Monte-Carlo. The size of this smearing, σ_{smear}, can be estimated by in-quadrature subtraction of the peak width in Monte-Carlo ($9.11\,\mathrm{MeV/c^2}$) from the width of peak in data ($10.19\,\mathrm{MeV/c^2}$), which yields $\sigma_{smear} = 4.57\,\mathrm{MeV/c^2}$.

We fit the self-reflection shape (Figure I-1 (b)) to a double Gaussian, to obtain the

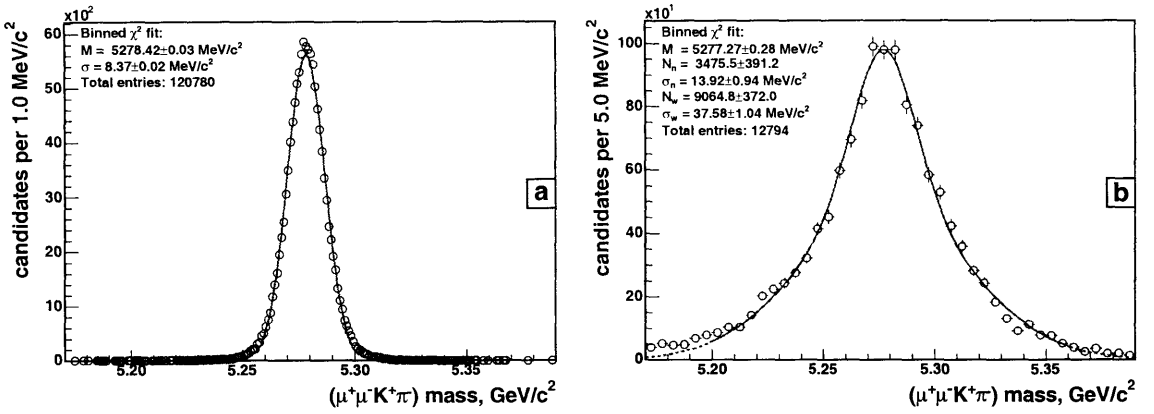


Figure I-1: Mass distribution of correctly reconstructed (a) and swapped (b) B_d candidates. Both distributions are obtained from the same Monte-Carlo sample.

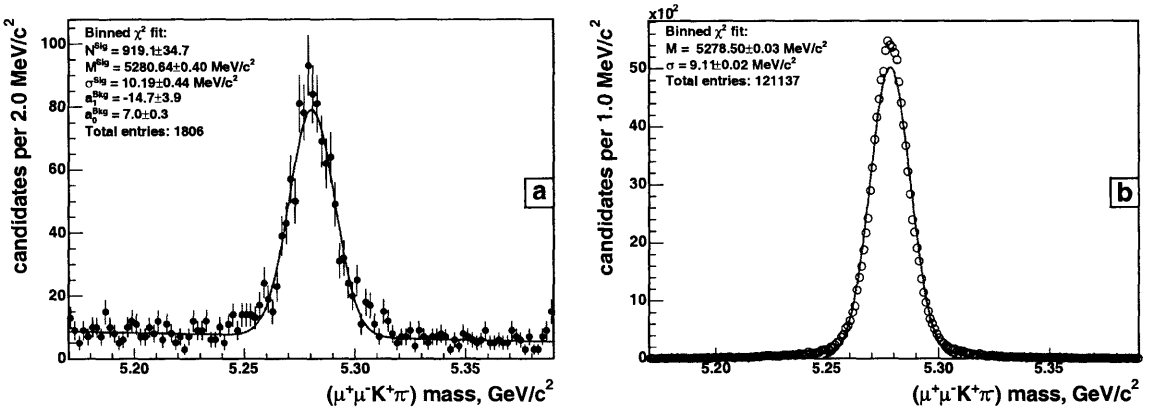


Figure I-2: Combined (correctly reconstructed and swapped) B_d mass candidates in: (a) data and (b) Monte-Carlo.

two widths, σ_n and σ_w, the offset ΔM_{sw} with respect to the mean of the unswapped Gaussian, M, and the ratio of the wide component to the total, f_w. The larger width, σ_w, is almost entirely governed by the kinematics, while the smaller width, σ_n, needs to be adjusted for the detector smearing not modeled in Monte-Carlo: $\sigma_n \to \sqrt{\sigma_n^2 + \sigma_{smear}^2}$.

Another issue of some importance is that the Gaussians, at least the wider one, are wide enough that they do not die out at the edges of the mass window, therefore a re-normalization is needed. The final shape that we use to describe the self-reflection in data is given by:

$$\begin{aligned} X^{obs}_{sig,sw}(m) =& (1 - f_w) \cdot G(m - (M + \Delta M_{sw}), \sigma_n)/N_n \\ &+ f_w \cdot G(m - (M + \Delta M_{sw}), \sigma_w)/N_w, \end{aligned} \tag{I.1}$$

where $N_{n,w}$ are given by $N_{n,w} = \frac{1}{2}\mathrm{Erf}\left(\frac{m-(M+\Delta M_{sw})}{\sqrt{2}\cdot\sigma_{n,w}}\right)\Big|_{m=5.17}^{m=5.39}$.

Finally, for the average lifetime measurement to properly construct the PDF of the

signal we need to know the fraction of the signal, f_{sw}, reconstructed in the swapped fashion. From the calculation based on the yields shown on plots (a) and (b) in Figure I-1 we get $f_{sw} = 9.56\%$.

The numerical values of all of the parameters are collected in Table I.1 and can also be deduced from the numbers on the plots illustrating this appendix.

Par.	Value	Unit
ΔM_{sw}	-1.23	MeV/c^2
σ_n	14.65	MeV/c^2
σ_w	37.58	MeV/c^2
f_w	72.29	%
f_{sw}	9.56	%

Table I.1: Summary of parameters describing the mass distribution of the B_d self-reflection when $K\pi$ masses are mis-assigned in the K^{*0} reconstruction.

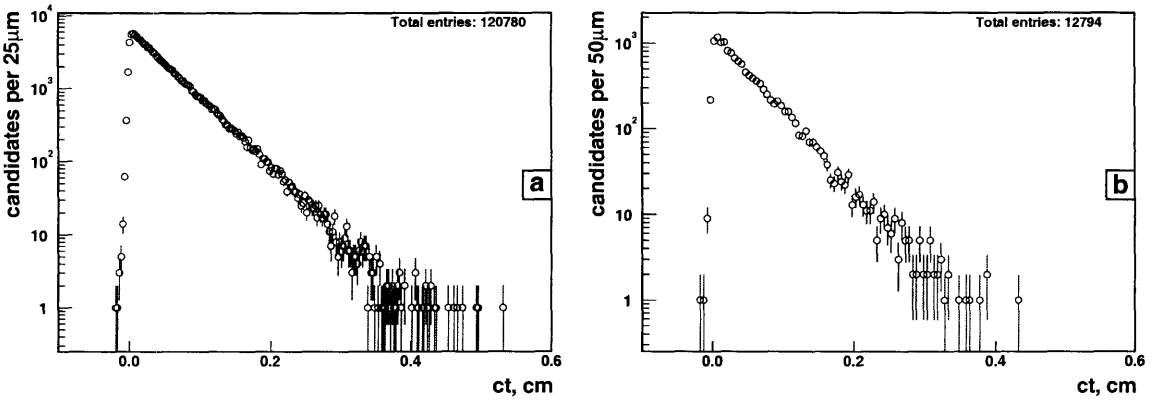


Figure I-3: PDL distribution of correctly reconstructed (a) and swapped (b) B_d candidates. Both distributions are obtained from the same Monte-Carlo sample.

To see how the lifetimes (Figure I-3) of the swapped and correctly reconstructed B_d candidates compare we fit an exponential convoluted with the Gaussian (event-by-event error) to both and compare the extracted lifetimes. The results are documented in Table I.2. The lifetimes agree within 0.2σ, therefore we use the same lifetime for

Sample	$c\tau$, μm	S_{ct}
Correct $K\pi$	466.12 ± 0.50	1.088 ± 0.005
Swapped $K\pi$	465.82 ± 1.43	1.056 ± 0.014

Table I.2: Summary lifetime fits on the correctly reconstructed B_d candidates and the self-reflection.

the swapped and correctly reconstructed B_d candidates when fitting the data. We neglect the difference in the error scale factor.

Appendix J

Fitting for angular decay amplitudes or time-integrated fractions

The time-dependent distribution for B_s decays is given by Equation 1.84, which we quote here for convenience:

$$\frac{d^4\mathcal{P}(\vec{\omega},t)}{d\vec{\omega}\,dt} \propto |A_0|^2 e^{-\Gamma_L t} f_1(\vec{\omega}) + |A_{\|}|^2 e^{-\Gamma_L t} f_2(\vec{\omega}) + |A_\perp|^2 e^{-\Gamma_H t} f_3(\vec{\omega}) + Re(A_0^* A_{\|}) e^{-\Gamma_L t} f_5(\vec{\omega}).$$

We want to use this as the PDF (before sculpting), and so need to normalize this function. The integral of the right-hand side of Equation 1.84 is $|A_0|^2\tau_L + |A_{\|}|^2\tau_L + |A_\perp|^2\tau_H$, since the integrals of $f_{1,2,3}(\vec{\omega})$ over the angles are 1, and the integral of $f_5(\vec{\omega})$ (and $f_{4,6}(\vec{\omega})$) over the angles are zero, independent of what the lifetimes are.

We have two possible ways to proceed. First, note that we can only measure the ratios of the various amplitudes $A_0, A_{\|}, A_\perp$, since we do not predict the number of events observed in our sample. We thus impose the condition that $|A_0|^2 + |A_{\|}|^2 + |A_\perp|^2 = 1$. In this case our normalized probability density is

$$\frac{d^4\mathcal{P}(\vec{\omega},t)}{d\vec{\omega}\,dt} = \frac{1}{(|A_0|^2 + |A_{\|}|^2)\tau_L + |A_\perp|^2\tau_H}\Big[\big[|A_0|^2 f_1(\vec{\omega}) + |A_{\|}|^2 f_2(\vec{\omega}) + Re(A_0^* A_{\|}) f_5(\vec{\omega})\big] e^{-t/\tau_L} + |A_\perp|^2 f_3(\vec{\omega}) e^{-t/\tau_H}\Big]. \quad \text{(J.1)}$$

This is the method we use to generate our final results.

Alternatively, we could introduce a set of variables C_α ($\alpha = 0, \|, \perp$):

$$C_0 = \frac{A_0\sqrt{\tau_L}}{\sqrt{|A_0|^2\tau_L + |A_{\|}|^2\tau_L + |A_\perp|^2\tau_H}}, \quad \text{(J.2)}$$

$$C_{\|} = \frac{A_{\|}\sqrt{\tau_L}}{\sqrt{|A_0|^2\tau_L + |A_{\|}|^2\tau_L + |A_\perp|^2\tau_H}}, \quad \text{(J.3)}$$

$$C_\perp = \frac{A_\perp\sqrt{\tau_H}}{\sqrt{|A_0|^2\tau_L + |A_\parallel|^2\tau_L + |A_\perp|^2\tau_H}}. \tag{J.4}$$

In terms of C_α the normalized probability density is

$$\begin{aligned}\frac{d^4\mathcal{P}(\vec{\omega},t)}{d\vec{\omega}\,dt} &= \left[|C_0|^2 f_1(\vec{\omega}) + |C_\parallel|^2 f_2(\vec{\omega}) + Re(C_0^* C_\parallel) f_5(\vec{\omega})\right]\frac{1}{\tau_L}e^{-t/\tau_L} \\ &+ |C_\perp|^2 f_3(\vec{\omega})\frac{1}{\tau_H}e^{-t/\tau_H},\end{aligned} \tag{J.5}$$

such that each exponential is separately normalized, but the overall normalization condition is still

$$|C_0|^2 + |C_\parallel|^2 + |C_\perp|^2 = 1. \tag{J.6}$$

In addition to having lower correlation coefficients to the lifetimes, the C_α give the time-integrated amounts of each angular component, *i.e.* $|C_\perp|^2$ is the fraction of CP-odd states in our sample. As such, they can be easily compared to the corresponding amplitudes reported in References [55, 106], which came from time-independent fits.

However, the total fraction of each component is not a physically transparent quantity. These fractions depend both on the amplitude to decay to that component, and the lifetime of the CP-even or odd eigenstate, which is determined by the totality of B_s decay modes available, of which the $J/\psi\phi$ mode is a small contributor. If the measurement is done in terms of C_α, one must convert them to the normalized amplitudes A_α.

Inverting Equation J.5 results in the following relations for the squares of the amplitudes:

$$|A_0|^2 = \frac{|C_0|^2}{1 - |C_\perp|^2(1 - \frac{\tau_L}{\tau_H})}, \tag{J.7}$$

$$|A_\parallel|^2 = \frac{|C_\parallel|^2}{1 - |C_\perp|^2(1 - \frac{\tau_L}{\tau_H})}, \tag{J.8}$$

$$|A_\perp|^2 = \frac{\tau_L}{\tau_H}\frac{|C_\perp|^2}{1 - |C_\perp|^2(1 - \frac{\tau_L}{\tau_H})}. \tag{J.9}$$

These should be equal to the squares of the normalized amplitudes obtained by fitting using the first method above. Note that, because the error on τ_H is large, these amplitudes will not be as well determined as the time-integrated C_α, which is expected: if we know the area of an exponential well, but don't know the lifetime well, the value at $t = 0$ will have a large uncertainty.

For the B_d, the two lifetimes are nearly the same, so Equations J.7–J.9 reduce to

$$|A_0|^2 = |C_0|^2, \tag{J.10}$$

$$|A_\parallel|^2 = |C_\parallel|^2, \tag{J.11}$$

$$|A_\perp|^2 = |C_\perp|^2. \tag{J.12}$$

The probability density used in the likelihood is, normalizing Equation 1.85:

$$\frac{d^4\mathcal{P}(\vec{\omega},t)}{d\vec{\omega}\,dt} = \frac{1}{|A_0|^2+|A_{\|}|^2+|A_{\perp}|^2}\Big[|A_0|^2 f_1(\vec{\omega}) + |A_{\|}|^2 f_2(\vec{\omega}) + |A_{\perp}|^2 f_3(\vec{\omega}) \pm Im(A_{\|}^* A_{\perp}) f_4(\vec{\omega}) + Re(A_0^* A_{\|}) f_5(\vec{\omega}) \pm Im(A_0^* A_{\perp}) f_6(\vec{\omega})\Big] \frac{1}{\tau_{B_d}} e^{-t/\tau_{B_d}}. \quad \text{(J.13)}$$

Assuming the condition $|A_0|^2 + |A_{\|}|^2 + |A_{\perp}|^2 = 1$ is already imposed, Equation J.13 becomes

$$\frac{d^4\mathcal{P}(\vec{\omega},t)}{d\vec{\omega}\,dt} = \Big[|A_0|^2 f_1(\vec{\omega}) + |A_{\|}|^2 f_2(\vec{\omega}) + |A_{\perp}|^2 f_3(\vec{\omega}) \pm Im(A_{\|}^* A_{\perp}) f_4(\vec{\omega}) + Re(A_0^* A_{\|}) f_5(\vec{\omega}) \pm Im(A_0^* A_{\perp}) f_6(\vec{\omega})\Big] \frac{1}{\tau_{B_d}} e^{-t/\tau_{B_d}}. \quad \text{(J.14)}$$

The normalized amplitudes A_α can be compared between B_d and B_s, which by flavor SU(3) are expected to be close. The comparison of the C_α for the B_d and B_s is not directly meaningful, as the different B_s lifetime components obscure the underlying physics. Figure J-1 illustrates this using the lifetimes and amplitudes obtained in this dissertation. For the B_d, since the P-odd and -even components have the same lifetime (to an excellent approximation), measuring the ratio of decays at any time, or integrated over time, gives the same result. However, for the B_s, the area under the two decay distributions is dependent both upon the decay amplitudes and the lifetimes.

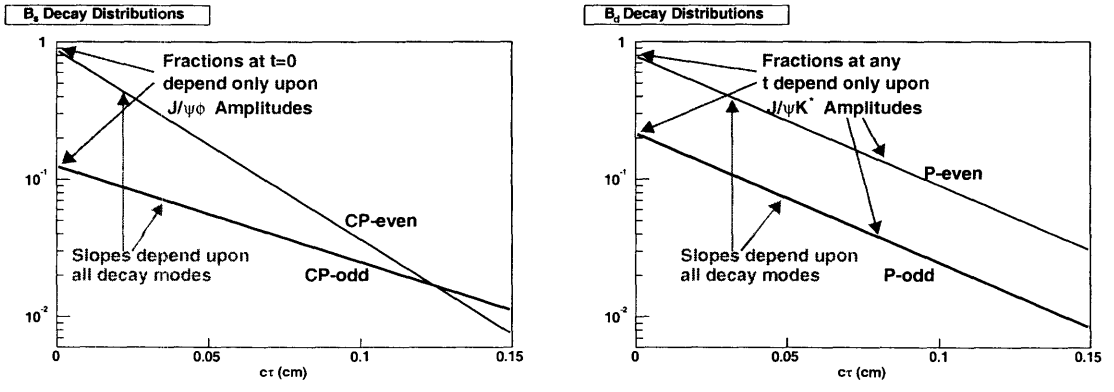


Figure J-1: PDL distributions for B_s, showing the effect of a non-zero $\Delta\Gamma$, and for B_d, where $\Delta\Gamma$ is assumed to be zero.

We perform both kinds of fits: using the the amplitudes A_α, and using the time-integrated fractions C_α. As expected, they give identical results (once the conversion given by Equations J.7–J.9 has been performed), but the first is more convenient for getting the uncertainties on the physical quantities, and the second for comparing to the previous time-independent fits and drawing fit projections (Appendix M).

Appendix K

Calculation of ξ_i

Section 3.2.4 introduces "partial normalization constants", ξ_i, which are to be computed by Monte-Carlo methods. Here we explain and document the details of such computation. We start from a simpler procedure of determining the ξ_i for the B_s, then continue by determining ξ_i^{un} and ξ_i^{sw} for the more complicated case of B_d mesons, and then conclude with a summary table and some remarks.

K.1 ξ_i for B_s

In order to determine $\xi_i^{B_s}$ we use Monte-Carlo Sample 3 described in Appendix C. This sample is generated uniformly in the phase-space, a nice feature of which is that the reconstructed distribution, say of the transversity angle ϕ, is in a sense the sculpting integrated over the rest of the variables. The sculpting of $\cos\theta$, ϕ and $\cos\psi$ distributions is shown in Figure 3-7.

Out of 40M generated about 516K events pass the analysis requirements. $\xi_i^{B_s}$ are determined by looping over all these events and applying the formula given by Equation 3.14. The numerical values of thus determined $\xi_i^{B_s}$ are collected in Table K.2 at the end of this appendix.

K.2 ξ_i^{un} and ξ_i^{sw} for B_d

The situation in the angular analysis of $B_d \to J/\psi K^{*0}$ decay is more complicated, because we need to determine two sets of ξ_i: $\{\xi_i^{un}\}$ and $\{\xi_i^{sw}\}$. The second set is the contribution from the events where $K\pi$ mass assignment is interchanged. To normalize the signal part of the likelihood we use $\xi_i = \xi_i^{un} + \xi_i^{sw}$, but the background (with efficiency defined by just $\epsilon^{un}(\vec{\omega})$) is normalized with just $\{\xi_i^{un}\}$. To know $\{\xi_i^{un}\}$ and $\{\xi_i^{sw}\}$ separately can also be useful in other situations, *e.g.* if one wants to calculate the total fraction of the swapped events.

To answer the challenge we generate 80M B_d decays into $J/\psi(\mu^+\mu^-)K^{*0}(K^+\pi^-)$ using `EvtGen` with `PHSP` decay model. After generated events are passed through the detector simulation and the analysis cuts/procedures we end up with about 725K

candidates (Sample 2 from Appendix C), about 10% of which are reconstructed with K and π mass assignments interchanged.

In the first step we need to determine the relative efficiency of the swapped and unswapped events $R(\vec{\omega}, \vec{\kappa})$ defined by Equation 3.28. Figure K-1 shows $r(\cos\theta)$, which is $R(\vec{\omega}, \vec{\kappa})$ integrated over the rest of the variables, as well as similarly defined $r(\phi)$, $r(\cos\psi)$, and $r(p_T^{B_d})$. $r(\cos\theta)$ and $r(\phi)$ are fit rather well by a flat line (slope of the

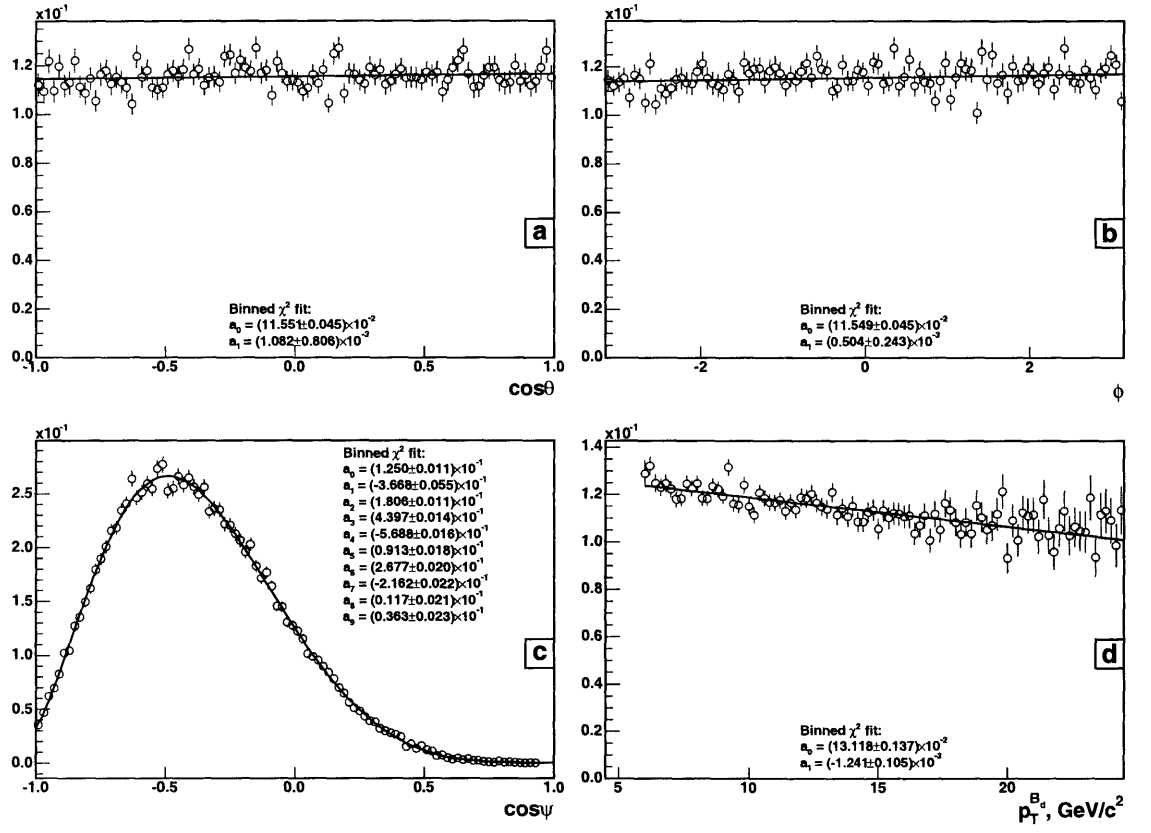


Figure K-1: Ratio of efficiencies for swapped and unswapped events in the B_d sample as a function of (a) $\cos\theta$, (b) ϕ, (c) $\cos\psi$, and (d) $p_T^{B_d}$.

linear fit is compatible with zero), which leaves only rather peculiar ways in which $R(\vec{\omega}, \vec{\kappa})$ can depend on $\cos\theta$ or ϕ. The dependence on $p_T^{B_d}$ (similar for $p_T^{K^{*0}}$, less pronounced for other kinematic variables we have looked at) appears to be linear, but rather weak at the same time. For the purposes of this analysis we choose to ignore the dependence of $R(\vec{\omega}, \vec{\kappa})$ on $\cos\theta$, ϕ or *any* kinematic quantity in $\vec{\kappa}$ and treat this approximation as a systematic effect (Section 3.4.5). With the above, $R(\vec{\omega}, \vec{\kappa})$ becomes $R(\cos\psi) = r(\cos\psi)$. We fit $R(\cos\psi)$ to a 9-degree polynomial to get the coefficients in Table K.1. Once $R(\cos\psi)$ is determined we proceed with calculating $\{\xi_i^{un}\}$ and $\{\xi_i^{sw}\}$ according to the prescription of Section 3.2.4. The resulting values are documented in Table K.2.

a_0	$1.25019 \cdot 10^{-1}$	a_3	$4.39667 \cdot 10^{-1}$	a_6	$2.67703 \cdot 10^{-1}$
a_1	$-3.66834 \cdot 10^{-1}$	a_4	$-5.68834 \cdot 10^{-1}$	a_7	$-2.16230 \cdot 10^{-1}$
a_2	$1.80602 \cdot 10^{-1}$	a_5	$9.12762 \cdot 10^{-2}$	a_8	$1.17344 \cdot 10^{-2}$
				a_9	$3.63272 \cdot 10^{-2}$

Table K.1: Coefficients of the polynomial describing $R(\cos\psi)$.

K.3 Summary

Table K.2 summarizes the numerical values of all ξ_i constants used in the analysis. Note the pattern $\xi_{1,2,3} \gg |\xi_5| > |\xi_{4,6}| \approx 0$. $\xi_{4,5,6}$ are expected to be small because

i	$\xi_i^{B_s}$	$\xi_i^{B_d,un}$	$\xi_i^{B_d,sw}$
1	$3.8096 \cdot 10^{-2}$	$3.4817 \cdot 10^{-2}$	$2.6987 \cdot 10^{-3}$
2	$4.0614 \cdot 10^{-2}$	$4.2330 \cdot 10^{-2}$	$5.6267 \cdot 10^{-3}$
3	$4.0662 \cdot 10^{-2}$	$4.2695 \cdot 10^{-2}$	$5.5877 \cdot 10^{-3}$
4	$-1.2821 \cdot 10^{-4}$ (0.0)	$4.3083 \cdot 10^{-5}$ (0.0)	$-5.7502 \cdot 10^{-6}$ (0.0)
5	$-7.4261 \cdot 10^{-5}$	$-7.2551 \cdot 10^{-4}$	$1.5917 \cdot 10^{-5}$
6	$4.3025 \cdot 10^{-5}$ (0.0)	$2.1007 \cdot 10^{-5}$ (0.0)	$4.4547 \cdot 10^{-6}$ (0.0)

Table K.2: Numerical values of the ξ_i factors used in the analysis.

$f_{4,5,6}(\vec{\omega})$ integrate to zero. The peculiarity of the sculpting may, however, pull them away from zero. Still $\xi_{4,6}$ end up being very close to zero because of the statistical cancellation between particles and anti-particles (we should have gotten zero had we used an infinite size Monte-Carlo sample). ξ_5 is small, but non-zero, because there is no such cancellation for this term. In doing the analysis we chose to fix all $\xi_{4,6}$ to zero as indicated in Table K.2.

Appendix L

Goodness of fit estimate

Fitting a model can be done using binned (histogrammed) or unbinned data. In the first case, the minimum χ^2 fit is most commonly used, which returns the fitted parameters, their statistical uncertainties as well as a measure of the goodness of fit. With unbinned data and the MLL fit, the fitted parameters with uncertainties are obtained, but no measure of goodness of fit is available. This remains an area in statistics where more effort is needed though some success is already achieved [111].

In our analysis we rely on unbinned maximum likelihood fits, but it is vital to get a quantitative estimate of how well fits describe the data. In order to proceed with such an estimate we have chosen to use a χ^2 test, which has a drawback of requiring us to bin the data, but has an advantage of easy interpretation. We have to bin the data anyway, if nothing else but for presentational purposes.

For concreteness we take the example of the average lifetime analysis. An extension of the procedure to the time-dependent angular analysis is discussed in Appendix M. There are two distributions that we fit our model to: mass and proper decay length. It is easier to estimate the quality of fit for the mass, so this is the one we start with.

Consider the PDF used in the fit (Equation 3.35 updated for $B_u \to J/\psi\pi^+$ contribution in case of B_u, and for the swapped component in the signal in case of B_d), integrated over ct. If one plugs in the fitted parameters and adds up the resulting functions for all candidates in the sample, a mass *fit projection curve* (or simply mass *projection*) is obtained. For a good fit the resulting projection should reasonably well approximate the corresponding data histogram (mass histogram in the current case). To see how well the two match, one can now create a *fit projection histogram* by integrating the fit projection curve over bins defined by the data histogram. In the mass histograms, at least those obtained from the data samples we deal with, any reasonable bin size guarantees that the number of entries in any bin is from several tens up to few hundreds, therefore one can reliably assume the uncertainty of bin content to be Gaussian. If the fit is even remotely good, each bin in the fit projection histogram

has at least several tens of entries too. We can then construct a χ^2_{test} variable:

$$\chi^2_{test} = \sum_{i=1}^{N_{bins}} \frac{(y_i^{data} - y_i^{proj})^2}{\sigma^2_{y_i^{proj}}} = \sum_{i=1}^{N_{bins}} \frac{(y_i^{data} - y_i^{proj})^2}{y_i^{proj}}, \qquad \text{(L.1)}$$

where y_i is the number of entries in the i^{th} bin of the corresponding histogram. We define a measure of goodness of fit as $Prob(\chi^2_{test}|N_{bins} - N_{par})$, where $N_{par} = 5$ is the number of fit parameters that are used to describe the mass distribution ($\{M, S_m, A, f_s\}$) plus one for the normalization automatically imposed by virtue of adding up PDFs each of which is normalized to unity.

We can estimate goodness of fit for the proper decay length in much the same way, except that there are now a number of bins that have very few or no entries and, consequently, non-Gaussian uncertainties. To avoid this problem and make applicable the procedure defined above we do the following:

a) bin the ct projection curve in bins of $10\,\mu$m (bin content is determined by integrating the projection between the lower and the upper edges of the bin),

b) step through all bins from step a) beginning with the leftmost, merging them until the total number of entries in the resulting "grand-bin" becomes 20 or larger. Once this is achieved we start a new "grand-bin",

c) as the result of step b) we have a certain unequal size binning which we now apply not only to the fit projection but to the data as well. We then construct the χ^2_{test} variable in the same way as we do for mass distribution (Equation L.1) and proceed with the test.

The measure of goodness of fit in this case is $Prob(\chi^2_{test}|N_{bins} - N_{par})$, where $N_{par} = 10$, assuming $\{f_s, c\tau, f_-, f_+, f_{++}, \lambda_-, \lambda_+, \lambda_{++}, S_\lambda\}$ is the set of parameters in the proper decay length part of the fit. The procedure is illustrated in Figure L-1.

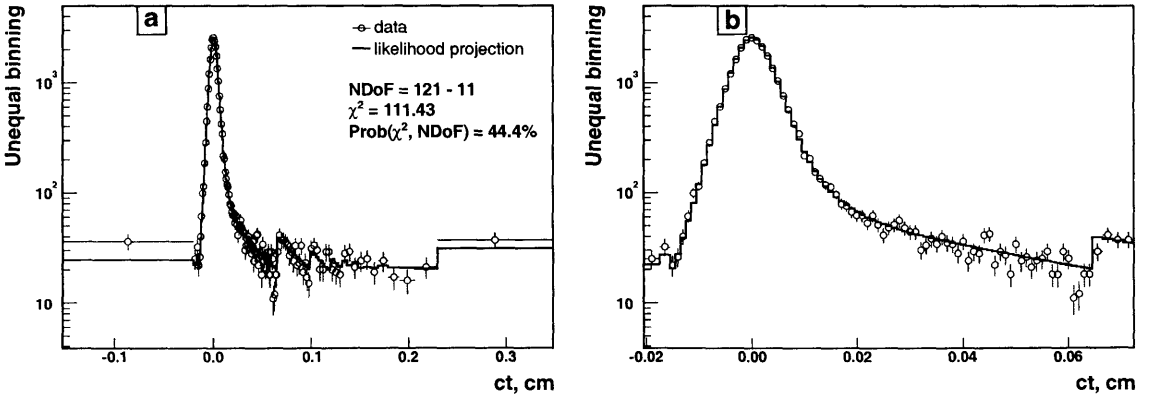


Figure L-1: (a) fit quality test for the proper decay length. Unequal binning according to the algorithm described in the text explains the shape of the likelihood projection and data. (b) zoom in on the region with the smallest bin size.

Appendix M

Projecting fit results on the angular distributions

The mechanism of avoiding the explicit inclusion of the efficiency and the kinematic density in the fit for the angular decay amplitudes makes it a non-trivial exercise to project the fit results on the angular distributions in data. To have such projections on the other hand is important for getting confidence in the fit results.

In order to make a projection for any of the transversity variables one needs the complete PDF which includes the product of the efficiency and the kinematic density, $V(\vec{\kappa})\epsilon(\vec{\omega},\vec{\kappa})$, whereas in the MLL fitting this product drops out. For the signal, $V(\vec{\kappa})\epsilon(\vec{\omega},\vec{\kappa})$ can be modeled using the Monte-Carlo simulation. For the background however, it is impossible to do, at least not without understanding all the sources and how much they contribute (and being able to simulate that with the Monte-Carlo simulation too!), as has been pointed out in Section 3.2.4.

The above limitation means that we cannot make projections involving the entire data sample, *i.e.* signal *and* background. At best we can make projections for the signal, but even that not without caveats:

- We can only isolate the signal *statistically* with methods like side-band subtraction, which has a bearing on the error bars in a histogram showing a particular signal distribution.
- Proper accounting of the sculpting needs to be done.

The analytical shape of the $\cos\theta$ projection is obtained by integrating the signal PDF (Equation J.1 for B_s, Equation J.14 for B_d) over ϕ, $\cos\psi$ and t. Similarly, ϕ and $\cos\psi$ projections are obtained by integrating the signal PDF over the other two transversity variables and time. Note, that we purposely referred to the signal PDF for B_s expressed in terms of time-integrated fractions (Appendix J), because the projection shapes calculated in terms of C_α are the same for B_d and B_s:

$$\frac{d\mathcal{P}(\cos\theta)}{d\cos\theta} = \frac{3}{8}\left(|C_0|^2+|C_{\|}|^2\right)(1+\cos^2\theta)+\frac{3}{4}|C_\perp|^2(1-\cos^2\theta) \quad \text{(M.1)}$$

$$\frac{d\mathcal{P}(\phi)}{d\phi} = \frac{3}{8\pi}\left[|C_0|^2\left(2-\frac{4}{3}\cos^2\phi\right)+|C_{\|}|^2\left(2-\frac{4}{3}\sin^2\phi\right)+\frac{4}{3}|C_\perp|^2\right] \quad \text{(M.2)}$$

$$\frac{d\mathcal{P}(\cos\psi)}{d\cos\psi} = \frac{3}{4}\left(1 - |C_0|^2\right) + \left(\frac{9}{4}|C_0|^2 - \frac{3}{4}\right)\cos^2\psi \qquad \text{(M.3)}$$

Let us now address the issue of sculpting. We have two choices: either sculpt the projection shapes given by Equations M.1–M.3, or un-sculpt the signal from data, before plotting them on top of each other. We choose to un-sculpt the data, for the technical reason that both data and the sculpting information exist in the form of histograms.

What are these sculpting histograms? As discussed in Section 3.2.4 and Appendix K, we can generate a Monte-Carlo sample of the decay of interest flat in all three transversity variables, in which case $\cos\theta$, ϕ, and $\cos\psi$ distributions of the reconstructed candidates give the shape of correspondingly $\cos\theta$, ϕ, and $\cos\psi$ sculpting. For the case of $B_s \to J/\psi\phi$ these distributions are shown in Figure 3-7. All we have to do to un-sculpt a transversity variable distribution in data is to divide it by the corresponding shape obtained from the Monte-Carlo.

Appendix N

Useful transformations

We collect here a set of useful relations for calculating and transforming between variables of interest. The set of variables is $\{\tau_L, \tau_H, \tau, \Gamma, \Delta\Gamma, \Delta\Gamma/\Gamma\}$. Any two can be chosen as the fit variables (with varying levels of correlation), from which the others can then be derived.

N.1 Variable definitions

Some of the definitions given here have already been introduced in Section 1.3 or are trivially obtained from those. However, we feel it is important to spell them out for the sake of completeness if nothing else. The definitions of Γ and $\Delta\Gamma$ are:

$$\Gamma \equiv (\Gamma_L + \Gamma_H)/2 \tag{N.1}$$

$$\Delta\Gamma \equiv \Gamma_L - \Gamma_H \tag{N.2}$$

so

$$\Gamma_L = \Gamma + \Delta\Gamma/2 \tag{N.3}$$

$$\Gamma_H = \Gamma - \Delta\Gamma/2 \tag{N.4}$$

The light and heavy lifetimes are:

$$\tau_L \equiv \frac{1}{\Gamma_L} \tag{N.5}$$

$$\tau_H \equiv \frac{1}{\Gamma_H} \tag{N.6}$$

Similarly τ is defined as the inverse of the mean width:

$$\tau \equiv \frac{1}{\Gamma} \tag{N.7}$$

Note that τ is a physically meaningful quantity, being the inverse of the total width. However, it is *not* the lifetime one would measure in a single-lifetime fit to a mixture

with equal amounts of heavy and light decays.

N.2 Relations

In terms of τ_L and τ_H:

$$\Delta\Gamma = \frac{1}{\tau_L} - \frac{1}{\tau_H} = \frac{\tau_H - \tau_L}{\tau_H \tau_L} \tag{N.8}$$

$$\Gamma = \frac{1}{2}\left(\frac{1}{\tau_L} + \frac{1}{\tau_H}\right) = \frac{\tau_H + \tau_L}{2\tau_H \tau_L} \tag{N.9}$$

$$\frac{\Delta\Gamma}{\Gamma} = \frac{2(\tau_H - \tau_L)}{\tau_H + \tau_L} \tag{N.10}$$

In terms of τ_L and $\Delta\Gamma$:

$$\frac{1}{\tau_H} = \frac{1}{\tau_L} - \Delta\Gamma = \frac{1 - \tau_L \Delta\Gamma}{\tau_L} \tag{N.11}$$

$$\Gamma = \frac{1}{\tau} = \frac{1}{\tau_L} - \frac{\Delta\Gamma}{2} = \frac{1 - \frac{1}{2}\tau_L \Delta\Gamma}{\tau_L} \tag{N.12}$$

$$\frac{\Delta\Gamma}{\Gamma} = \frac{\tau_L \Delta\Gamma}{1 - \frac{1}{2}\tau_L \Delta\Gamma} \tag{N.13}$$

In terms of τ_L and $\Delta\Gamma/\Gamma$:

$$\Gamma_L = \Gamma\left(1 + \frac{1}{2}\frac{\Delta\Gamma}{\Gamma}\right) = \frac{1}{\tau_L} \tag{N.14}$$

$$\Gamma_H = \Gamma\left(1 - \frac{1}{2}\frac{\Delta\Gamma}{\Gamma}\right) = \frac{1}{\tau_H} \tag{N.15}$$

so

$$\Gamma = \frac{1}{(1 + \frac{1}{2}\frac{\Delta\Gamma}{\Gamma})\tau_L} \tag{N.16}$$

$$= \frac{1}{(1 - \frac{1}{2}\frac{\Delta\Gamma}{\Gamma})\tau_H} \tag{N.17}$$

$$\tau_H = \left(\frac{1 + \frac{1}{2}\frac{\Delta\Gamma}{\Gamma}}{1 - \frac{1}{2}\frac{\Delta\Gamma}{\Gamma}}\right)\tau_L \tag{N.18}$$

$$\Delta\Gamma = \frac{\Delta\Gamma}{\Gamma}\Gamma = \frac{\frac{\Delta\Gamma}{\Gamma}}{(1 + \frac{1}{2}\frac{\Delta\Gamma}{\Gamma})\tau_L} \tag{N.19}$$

In terms of τ and $\Delta\Gamma/\Gamma$:

$$\tau_L = \frac{\tau}{(1 + \frac{1}{2}\frac{\Delta\Gamma}{\Gamma})} \tag{N.20}$$

$$\tau_H = \frac{\tau}{(1-\frac{1}{2}\frac{\Delta\Gamma}{\Gamma})} \tag{N.21}$$

$$\Delta\Gamma = \frac{1}{\tau}\frac{\Delta\Gamma}{\Gamma} \tag{N.22}$$

In terms of τ and $\Delta\Gamma$:

$$\tau_L = \frac{1}{1/\tau+\frac{1}{2}\Delta\Gamma} = \frac{\tau}{1+\frac{1}{2}\tau\Delta\Gamma} \tag{N.23}$$

$$\tau_H = \frac{1}{1/\tau-\frac{1}{2}\Delta\Gamma} = \frac{\tau}{1-\frac{1}{2}\tau\Delta\Gamma} \tag{N.24}$$

$$\frac{\Delta\Gamma}{\Gamma} = \tau\Delta\Gamma \tag{N.25}$$

In terms of τ and τ_L:

$$\frac{1}{\tau_H} = \frac{2}{\tau}-\frac{1}{\tau_L} = \frac{2\tau_L-\tau}{\tau\tau_L} \tag{N.26}$$

$$\Delta\Gamma = \frac{1}{\tau_L}-\frac{1}{\tau_H} = 2(\frac{1}{\tau}-\frac{1}{\tau_L}) = 2\frac{\tau-\tau_L}{\tau\tau_L} = 2(\frac{\tau}{\tau_L}-1)\frac{1}{\tau} \tag{N.27}$$

$$\frac{\Delta\Gamma}{\Gamma} = 2(\frac{\tau}{\tau_L}-1) \tag{N.28}$$

$$\tag{N.29}$$

N.3 Summary

$$\tau_L \equiv \frac{1}{\Gamma_L} = \frac{1}{\Gamma+\frac{1}{2}\Delta\Gamma} = \frac{1}{\Gamma}\frac{1}{1+\frac{1}{2}\frac{\Delta\Gamma}{\Gamma}} = \frac{\tau}{1+\frac{1}{2}\frac{\Delta\Gamma}{\Gamma}} = \frac{\tau}{1+\frac{1}{2}\tau\Delta\Gamma}$$

$$\tau_H \equiv \frac{1}{\Gamma_H} = \frac{1}{\Gamma-\frac{1}{2}\Delta\Gamma} = \frac{1}{\Gamma}\frac{1}{1-\frac{1}{2}\frac{\Delta\Gamma}{\Gamma}} = \frac{\tau}{1-\frac{1}{2}\frac{\Delta\Gamma}{\Gamma}} = \frac{\tau}{1-\frac{1}{2}\tau\Delta\Gamma} = \frac{\tau_L}{1-\tau_L\Delta\Gamma} =$$

$$= \frac{\tau\tau_L}{2\tau_L-\tau} = \frac{\tau}{2-\frac{\tau}{\tau_L}} = \frac{1+\frac{1}{2}\frac{\Delta\Gamma}{\Gamma}}{1-\frac{1}{2}\frac{\Delta\Gamma}{\Gamma}}\tau_L$$

$$\tau \equiv \frac{1}{\Gamma} = 2\frac{\tau_L\tau_H}{\tau_H+\tau_L} = \frac{\tau_L}{1-\frac{1}{2}\tau_L\Delta\Gamma} = (1+\frac{1}{2}\frac{\Delta\Gamma}{\Gamma})\tau_L$$

$$\Gamma \equiv \frac{1}{2}(\Gamma_L+\Gamma_H) = \frac{\tau_H+\tau_L}{2\tau_L\tau_H} = \frac{1-\frac{1}{2}\tau_L\Delta\Gamma}{\tau_L} = \frac{1}{(1+\frac{1}{2}\frac{\Delta\Gamma}{\Gamma})\tau_L}$$

$$\Delta\Gamma \equiv \Gamma_L-\Gamma_H = \frac{\tau_H-\tau_L}{\tau_L\tau_H} = \frac{1}{\tau}\frac{\Delta\Gamma}{\Gamma} = 2\frac{\tau-\tau_L}{\tau\tau_L} = \frac{2}{\tau}(\frac{\tau}{\tau_L}-1) = \frac{\frac{\Delta\Gamma}{\Gamma}}{(1+\frac{1}{2}\frac{\Delta\Gamma}{\Gamma})\tau_L}$$

$$\frac{\Delta\Gamma}{\Gamma} = 2\frac{\Gamma_L-\Gamma_H}{\Gamma_L+\Gamma_H} = 2\frac{\tau_H-\tau_L}{\tau_H+\tau_L} = \frac{\tau_L\Delta\Gamma}{1-\frac{1}{2}\tau_L\Delta\Gamma} = \tau\Delta\Gamma$$

Bibliography

[1] E. Fermi, **Versuch einer Theorie der β-Strahlen. I**, Z. Phys. **88** 161-177, (1934).

[2] T. Lee and C. Yang, **Question of parity conservation in weak interactions**, Phys. Rev. **104** 254-258, (1956).

[3] C. Wu *et. al.*, **Experimental test of parity conservation in β decay**, Phys. Rev. **105** 1413-1415, (1957).

[4] J. Sakurai, **Mass reversal and weak Interactions**, Nuovo Cim. **7** 649-660, (1958).

[5] S. Glashow, **Partial-symmetries of weak interactions**, Nucl. Phys. **22** 579-588, (1961).

[6] S. Weinberg, **A model of leptons**, Phys. Rev. Lett. **19** 1264-1266, (1967).

[7] A. Salam, **Weak and electromagnetic interactions**, Proceedings of the Eights Nobel Symposium, Ed. N. Svartholm (Almqvist and Wiksells, Stockholm) 367-377, (1968).

[8] S. Weinberg, **Physical processes in a convergent theory of the weak and electromagnetic interactions**, Phys. Rev. Lett. **27** 1688-1691, (1971).

[9] P. Higgs, **Broken symmetries and the masses of gauge bosons**, Phys. Rev. Lett. **13** 508-509, (1964).

[10] P. Higgs, **Spontaneous symmetry breakdown without massless bosons**, Phys. Rev. **145** 1156-1163, (1966).

[11] M. Gell-Mann, **A schematic model of barions and mesons**, Phys. Lett. **8** 214-215, (1964).

[12] G. Zweig, **An SU(3) model for strong interaction symmetry and its breaking**, CERN-TH-412, (1964).

[13] N. Cabibbo, **Unitary symmetry and leptonic decays**, Phys. Rev. Lett. **10** 531-533, (1963).

[14] S. Glashow, J. Iliopoulos, L. Miani, **Weak interactions with lepton-hadron symmetry**, Phys. Rev. **D2** 1285-1292, (1970).

[15] E. Bloom *et. al.*, **High-energy inelastic $e-p$ scattering at 6° and 10°**, Phys. Rev. Lett. **23**, 930-934 (1969).

[16] M. Kobayashi and T. Maskawa, **CP violation in the renormalizable theory of weak interaction**, Progr. Theor. Phys. **49**, 652-657 (1973).

[17] J. Christenson, J. Cronin, V. Fitch, and R. Turlay **Evidence for the 2π decay of the K_2^0 meson**, Phys. Rev. Lett. **13**, 138-140 (1964).

[18] J. Aubert *et. al.*, **Experimental observation of a heavy particle J**, Phys. Rev. Lett. **33**, 1404-1406 (1974).

[19] J. Augustin *et. al.*, **Discovery of a narrow resonance in e^+e^- annihilation**, Phys. Rev. Lett. **33**, 1406-1408 (1974).

[20] S. Herb *et. al.*, **Observation of dimuon resonance at 9.5 GeV in 400- GeV proton-nucleus collisions**, Phys. Rev. Lett. **39**, 252-255 (1977).

[21] C. Rubbia *et. al.*, **Experimental observation of the intermediate vector bosons W^+, W^-, and Z^0**, Rev. Mod. Phys. **57**, 699-722 (1985).

[22] F. Abe *et. al.*, **Observation of top quark production in $p\bar{p}$ collisions with the Collider Detector at Fermilab**, Phys. Rev. Lett. **74**, 2626-2631 (1995).

[23] S. Abachi *et. al.*, **Observation of the top quark**, Phys. Rev. Lett. **74**, 2632-2637 (1995).

[24] D. Bardin, **Field theory and the Standard Model**, *http://doc.cern.ch/yellowrep/2000/2000-007/p1.pdf.*

[25] K. Hagiwara *et. al.*, **Review of particle physics**, Phys. Rev. **D66**, 010001 (2002) and 2003 off-year partial update for the 2004 edition available on the PDG WWW pages (*http://pdg.lbl.gov/*).

[26] L. Wolfenstein, **Parametrization of the Kobayashi-Maskawa Matrix**, Phys. Rev. Lett. **51** 1945-1947 (1983).

[27] C. Jarlskog, **Commutator of the quark mass matrices in the standard electroweak model and a measure of maximal CP nonconservation**, Phys. Rev. Lett. **55** 1039-1042 (1985).

[28] C. Kim and A. Martin, **On the determination of V_{ub} and V_{cb} from semileptonic B decays**, Phys. Lett. **B225** 186-190 (1989).

[29] I. Bigi, **The QCD perspective on lifetimes of heavy-flavour hadrons**, UND-HEP-95-BIG02 (1995).

[30] K. Anikeev *et. al.*, **B physics at the Tevatron Run II and beyond**, FERMILAB-Pub-01,197 (2001).

[31] N. Uraltsev, **Heavy Quark Expansion in beauty and its decays**, UND-HEP-98-BIG1 (1998).

[32] K. Wilson, **Non-lagrangian models of current algebra**, Phys. Rev. **179** 1499-1512 (1969).

[33] I. Bigi, **The lifetimes of heavy flavour hadrons — a case study in quark-hadron duality**, UND-HEP-99-BIG07 (1999).

[34] M. Beneke, G. Buchalla, C. Greub, A. Lenz, and U. Nierste, **The $B^+ - B_d^0$ lifetime difference beyond leading logarithms**, arXiv:hep-ph/0202106 (2002).

[35] M. Beneke, G. Buchalla, C. Greub, A. Lenz and U. Nierste, **Next-to-leading order QCD corrections to the lifetime difference of B_s mesons**, Phys Lett. **B459** 631-640 (1999),
U. Nierste, private communication.

[36] E. Franco, V. Lubicz, F. Mescia, and C. Tarantino, **Lifetime ratios of beauty hadrons at the next-to-leading order in QCD**, arXiv:hep-ph/0203089 (2002).

[37] V. Weisskopf and E. Wigner, **Berechnung der nat urlichen Linienbreite auf Grund der Diracschen Lichttheorie**, Z. Phys. **63** 54-73 (1930).

[38] M. Beneke, G. Buchalla, and I. Dunietz, **Width difference in the $B_s - \overline{B}_s$ system**, Phys. Rev. **D54** 4419-4431 (1996).

[39] A. Buras, W. Slominski, and H. Steger, **$B^0 - \overline{B^0}$ mixing, CP violation and the B-meson decay**, Nucl. Phys. **B245** 369-398 (1984).

[40] I. Bigi *et. al.*, **CP violation**, ed. C. Jarlskog (Singapore World Scientific) (1989).

[41] I. Dunietz, **$B_s - \overline{B}_s$ mixing, CP violation and extraction of CKM phases from untagged B_s data samples**, Phys. Rev. **D52** 3048-3064 (1995).

[42] R. Aleksan, A. Le Yaouanc, L. Oliver, O. Pène, and J.-C. Raynal, **Estimation of $\Delta\Gamma$ for the $B_s - \overline{B}_s$ system. Exclusive decays and the parton model**, Phys. Lett. **B316** 567-577 (1993).

[43] M. Beneke, G. Buchalla, A. Lenz, and U. Nierste, **CP asymmetry in flavour-specific B decays beyond leading logarithms**, Phys. Lett. **B576** 173-183 (2003).

[44] The BaBar Collaboration, **Limits on the lifetime difference of neutral B mesons and on CP, T, and CPT violation in $B^0 - \overline{B^0}$ mixing**, SLAC-PUB-9696 (2003).

[45] I. Dunietz, R. Fleischer, and U. Nierste, **In pursuit of new physics with B_s decays**, Phys. Rev. **D63** 114015 (2001).

[46] A. Dighe, I. Dunietz, H. Lipkin, and J. Rosner, **Angular distributions and lifetime differences in $B_s \to J/\psi\phi$ decays**, Phys. Lett. **B369** 144-150 (1996).

[47] A. Dighe, I. Dunietz, and R. Fleischer, **Extracting CKM phases and $B_s - \overline{B}_s$ mixing parameters from angular distributions of non-leptonic B decays**, Eur. Phys. J. **C6** 647 (1999).

[48] A. Dighe *et. al.*, **Measurement of the lifetime difference of B_d mesons: possible and worthwhile?**, Nucl. Phys. **B624** 377-404 (2002).

[49] A. Korn, C. Paus, and D. Starr, **Update on the measurement of B meson masses in the exclusive J/ψ channels**, CDF Note 6426, (2003).

[50] The CDF Collaboration, **Measurement of B hadron lifetimes using J/ψ final states at CDF**, Phys. Rev. **D57** 5382-5401 (1998).

[51] M. Wirbel, B. Stech, and M. Bauer, **Exclusive semileptonic decays of heavy mesons**, Z. Phys. **C29** 637 (1985). **Exclusive nonleptonic decays of D, D_s, and B mesons**, Z. Phys. **C34** 103 (1987).

[52] N. Isgur, D. Scora, B. Grinstein, and M. Wise, **Semileptonic B and D decays in the quark model**, Phys. Rev. **D39** 799 (1989).

[53] R. Aleksan, A. Le Yaouanc, L. Oliver, O. Pène, and J.-C. Raynal, **Critical analysis of theoretical estimates for B-to-light-meson form factors and $B \to J/\psi K(K^*)$ data using factorization**, Phys. Rev. **D51** 6235 (1995).

[54] CLEO Collaboration, **Measurement of the decay amplitudes and branching fractions of $B \to J/\psi K^*$ and $B \to J/\psi K$ decays**, Phys. Rev. Lett. **79** 4533-4537 (1997).

[55] CDF Collaboration, **Measurement of the decay amplitudes of $B^0 \to J/\psi K^{*0}$ and $B_s^0 \to J/\psi\phi$ decays**, Phys. Rev. Lett. **85** 4668 (2000).

[56] BaBar Collaboration, **Measurement of the $B \to J/\psi K^*(892)$ decay amplitudes**, Phys. Rev. Lett. **87** 241801 (2001).

[57] Belle Collaboration, **Measurement of the branching fractions and decay amplitudes in $B \to J/\psi K^*$ decays**, Phys. Lett. **B538** 11-20 (2002).

[58] K. Hartkorn, and H.-G. Moser, **A new method of measuring $\frac{\Delta\Gamma}{\Gamma}$ in the $B_s - \overline{B}_s$ system**, Eur. Phys. J. **C8** 381 (1999).

[59] ALEPH Collaboration, **A study of the decay width difference in the $B_s - \overline{B}_s$ system using $\phi\phi$ correlations**, Phys. Lett. **B486** 286-299 (2000).

[60] ALEPH, CDF, DELPHI, L3, OPAL, and SLD Collaborations, **Combined results on b-hadron production rates, lifetimes, oscillations and semileptonic decays**, hep-ex/0009052 (2000).

[61] CDF Collaboration, **The CDF II detector. Technical design report**, FERMILAB-Pub-96/390-E, (1996).

[62] Fermilab Beams Division, **Run II handbook**, *http://www-bdnew.fnal.gov/pbar/run2b/Documents/RunII_handbook.pdf*, (1999).

[63] Fermilab Beams Division, **Fermilab Linac upgrade. Conceptual design**, FERMILAB-LU-Conceptual Design *http://www-lib.fnal.gov/archive/linac/FERMILAB-LU-ConceptualDesign.pdf*, (1989).

[64] Fermilab Beams Division, **Booster rookie book**, *http://www-bd.fnal.gov/proton/booster/rookie/realrookiebook.html*

[65] Fermilab Beams Division, **Accelerator concepts rookie book**, *http://www-bdnew.fnal.gov/operations/rookie_books/concepts/conceptstoc.html*

[66] Fermilab Beams Division, **The antiproton source rookie book**, *http://www-bdnew.fnal.gov/pbar/documents/PBAR_Rookie_Book.PDF*, (1999).

[67] Fermilab Beams Division, **Recycler design report**, FERMILAB-TM-1991 *http://www-lib.fnal.gov/archive/1997/tm/TM-1991.html*, (1996).

[68] S. van der Meer, **Stochastic damping of betatron oscillations in the ISR**, CERN/ISR-PO/72-31, (1994).

[69] S. Nagaitsev, **Status of the Fermilab electron cooling project**, EPAC2002 Paris (2002).

[70] H. Montgomery, **Physics with the Main Injector** (DPF99 Los Angeles), *http://www.dpf99.library.ucla.edu/browsedpf99.html*, (1999).

[71] E. Prebys, **The Fermilab accelerator complex: status and challenges** (talk at the Tevatron University), *http://www.fnal.gov/orgs/utev/talk-slides/Eric.ps*, (2002).

[72] M. Church, **Tevatron Run II performance and plans**, FERMILAB-Conf-02/142-E, (2002).

[73] J. Bjorken, and E. Paschos, **Inelastic $e - p$ and $\gamma - p$ scattering and the structure of the nucleon**, Phys. Rev. **175** (1975).

[74] C. Blocker, *http://cdfkits.fnal.gov/CdfCode/source/VertexAlg/doc/*

[75] J. Marriner, **Secondary vertex fit with mass and pointing constraints (CTVMFT)**, CDF Note 1996, (1993).

[76] A. Mukherjee, **CTC and VTX tracking**, CDF Note 5490, (2000).

[77] F. Bedeschi, **An inexpensive method for global track fitting and track segment match**, CDF Note 0371, (1987).

[78] P. Azzi, G. Busetto, P. Gatti, A. Ribbon, **Histogram tracking in the COT**, CDF Note 5562, (2001).

[79] D. Amidei *et. al.*, **The silicon vertex detector of the Collider Detector at Fermilab**, NIMPR A350 73, (1994).

[80] W.-M. Yao, and K. Bloom, **"Outside-In" silicon tracking at CDF**, CDF Note 5991, (2002).

[81] G. Ascoli *et. al.*, **The CDF central muon detector**, NIM, Volume A, 268, p.63-67, (1988).

[82] The CDF Collaboration, **The 1992 CDF muon system upgrade**, CDF Note 2858, (1994).

[83] L. Balka, **The CDF central electromagnetic calorimeter**, NIMPR A267 272, (1988).

F. Abe *et. al.*, Phys. Rev. **D44** 29, (1991).

F. Abe *et. al.*, Phys. Rev. **D48** 2998, (1993).

F. Abe *et. al.*, Phys. Rev. Lett. **73** 2662, (1994).

J. Strait, **Design considerations for a preshower detector for the CDF end plugs**, CDF Note 2097, (1993).

S. Bertolucci, **The CDF central and endwall hadron calorimeter**, NIMPR A267 301, (1988).

[84] The CDF II Collaboration, Proposal for Enhancement of the CDF II Detector: **An Inner Silicon Layer and a Time of Flight detector (P-909)**, *http://www-cdf.fnal.gov/upgrades/PAC/p909.ps*, (1998)
Update to Proposal P-909: **Physics performance of the CDF II with an Inner Silicon Layer and a Time of Flight detector**, *http://www-cdf.fnal.gov/upgrades/btb_update_jan99.ps*, (1999).

[85] S. Klimenko *et. al.*, **The CDF Run II luminosity monitor**, CDF Note 4330, (1998).

S. Klimenko *et. al.*, **CLC production and mechanical assembly**, CDF Note 5119, (1999).

D. Acosta *et. al.*, **A first look at the CLC luminosity measurements**, CDF Note 6052, (2002).

[86] The CDF II Collaboration, **CDF Run II trigger table and datasets plan**, CDF Note 4718, (2001).

[87] H. Frisch *et. al.*, **Conceptual design of a deadtimeless trigger for the CDF trigger upgrade**, CDF Note 2038, (1994).

[88] E. Thomson *et. al.*, **Online track processor for the CDF upgrade**, IEEE Transactions on Nuclear Science, Vol. 49, No. 3, (2002).

[89] M. Paulini, and B. Wicklund, **Summary of proposal for B physics triggers in Run II**, CDF Note 5483, (2000).

K. Anikeev *et. al.*, **Level-1 di-muon trigger study (CMU-CMU)**, CDF Note 5769, (2002).

M. Bishai *et. al.*, **Run II di-muon trigger optimization and efficiency measurement**, CDF Note 6004, (2002).

[90] K. Anikeev *et. al.*, **Event Builder and Level-3 for aces**, CDF Note 5793, (2001).

[91] S. Giagu, **ROOT macros and instructions for accessing DB good run bits**,
http://www-cdf.fnal.gov/internal/physics/bottom/validation/rootmacro.html

[92] CDF Collaboration, **Measurement of the J/ψ meson and b quark production cross sections in $p\bar{p}$ collisions at $\sqrt{s} = 1960$ GeV**, to be published in Phys. Rev. **D**.

[93] R. Brun, and F. Rademakers, **ROOT – an object oriented data analysis framework**, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86; see also *http://root.cern.ch/*.

[94] K. Anikeev *et. al.*, **Calibration of energy loss and magnetic field using J/ψ events in Run II**, CDF Note 5958, (2002).

M. Campanelli, and E. Gerchtein, **Calibration of the momentum scale for Kalman refitter using J/ψ events**, CDF Note 6905, (2004).

[95] C. Hays, **Default tracks in 5.3.1**, Talk in the BPAK meeting. 02/27/2004.

[96] CDF Collaboration, **Measurement of B hadron lifetimes using J/ψ final states at CDF**, Phys. Rev. **D57** 5382 (1998).

[97] J. Lee, I. Cho, I. Yu, and T. Miao **Measurement of $Br(B^+ \to J/\psi\pi^+)/Br(B^+ \to J/\psi K^+)$**, CDF Note 6967, (2004).

[98] J. Orear, **Notes on statistics for physicists**, UCRL-8417, (1958).

[99] In order to integrate a function over a complicated domain $\boldsymbol{D}$, Monte Carlo integration picks random points over some simple domain $\boldsymbol{D'}$ which is a superset of $\boldsymbol{D}$, checks whether each point is within $\boldsymbol{D}$, and estimates the area of $\boldsymbol{D}$ (volume, n-dimensional content, *etc.*) as the area of $\boldsymbol{D'}$ multiplied by the fraction of points falling within $\boldsymbol{D'}$ (from *http://mathworld.wolfram.com/MonteCarloIntegration.html*).
J. Hammersley, **Monte-Carlo methods for solving multivariable problems**, Ann. New York Acad. Sci. **86**, 844-874, (1960).

[100] Instead of maximizing the likelihood $\mathcal{L}$ we actually minimize $-2\log\mathcal{L}$, using the HEP standard minimization package `MINUIT`. Factor of 2 in front of the log-likelihood is needed for correct estimates of the uncertainties. `MINOS` is a subpackage of `MINUIT` to estimate the uncertainties correctly even if the shape of the log-likelihood at the minimum is not parabolic.
F. James, **MINUIT. Function minimization and error analysis**, CERN Program Library Long Write-up D506; see also
http://consult.cern.ch/writeup/minuit/

[101] In consultation with Ray Culbertson (the CDF Alignment Team leader at the time) whom we thank very much for his time and effort.

[102] K. Anikeev, G. Bauer, and C. Paus, **B Meson lifetime measurements using exclusively reconstructed decays $B \to J/\psi X$ in Run II data**, CDF Note 6266, (2003).

[103] J. Boudreau *et. al.*, **Measurement of exclusive B lifetimes in the modes: $B^+ \to J/\psi K^+$, $B^+ \to J/\psi K^{*+}$, $B^0 \to J/\psi K^{*0}$, and $B^0 \to J/\psi K^0_S$ and measurement of the B^+/B^0 lifetime ratio**, CDF Note 6387, (2004).

[104] R. Barlow, **Systematic errors: facts and fiction**, Proc. of the conf. on advanced statistical techniques in particle physics, Ed. M. Whalley and L. Lyons (Durham, IPPP 02/39), 134-144, (2002); see also
http://www.ippp.dur.ac.uk/Workshops/02/statistics/proceedings/barlow.ps

[105] K. Anikeev, and C. Paus, **New measurement of lifetimes of B mesons in the exclusively reconstructed decays $B^+ \to J/\psi K^+$, $B^0 \to J/\psi K^{*0}$, $B_s \to J/\psi\phi$**, CDF Note 6551, (2004).

[106] C. Gay, K. Li, M. Schmidt, **Measurement of polarization amplitudes of $B_s \to J/\psi\phi$ and $B_d \to J/\psi K^{*0}$**, CDF Note 6960, (2004).

[107] P. Sphicas, **A $b\bar{b}$ Monte-Carlo generator**, CDF Note 2655, (1994).

K. Anikeev, P. Murat, C. Paus, **Description of Bgenerator II**, CDF Note 5092, (1999).

[108] A. Ryd *et. al.*, **EvtGen. A Monte-Carlo generator for B physics**, *http://www.slac.stanford.edu/ lange/EvtGen/*

W. Bell *et. al.*, **User guide for EvtGen at CDF**, CDF Note 5618, (2003).

[109] R. Brun and F. Carminati, **GEANT detector description and simulation tool**, CERN Program Library Long Write-up W5013; see also *http://consult.cern.ch/writeup/geant/*

[110] S. D'Auria *et. al.*, **CDF B Monte-Carlo**, *http://www-cdf.fnal.gov/internal/physics/bottom/b-montecarlo/*

[111] R. Raja, **A measure of the goodness of fit in unbinned likelihood fits**, FNAL-PUB-02/152-E (2002).

[112] Heavy Flavor Averaging Group, **Results for the winter 2004 conferences**, *http://www.slac.stanford.edu/xorg/hfag/osc/winter_2004/index.html#DG*